

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ –
УЧЕБНО-НАУЧНО-ПРОИЗВОДСТВЕННЫЙ КОМПЛЕКС»

На правах рукописи

ЛУКЪЯНОВ ПАВЕЛ ВАДИМОВИЧ

**АВТОМАТИЗАЦИЯ ПРОЦЕССА АДАПТАЦИИ В СИСТЕМЕ
АДМИНИСТРАТИВНОГО МОНИТОРИНГА**

05.13.06 – Автоматизация и управление технологическими
процессами и производствами (промышленность)

Диссертация
на соискание ученой степени
кандидата технических наук

Научный руководитель:
доктор технических наук, профессор
КОНСТАНТИНОВ ИГОРЬ СЕРГЕЕВИЧ

Орел, 2014

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	5
1 Анализ средств и методов организации административного мониторинга.....	13
1.1 Сущность административного мониторинга и его роль в принятии управленческих решений.....	13
1.1.1 Место автоматизированной системы административного мониторинга в системах управления.....	13
1.1.2 Основные понятия и положения организации мониторинга.....	19
1.1.3 Существующие виды систем мониторинга.....	22
1.1.4 Специфика административного мониторинга.....	25
1.2 Проблемы процесса адаптации и пути их решения.....	27
1.2.1 Отдельные проблемы применения больших сложных АСУ.....	27
1.2.2 Организация процесса адаптации системы административного мониторинга.....	28
1.2.3 Задачи получения, анализа и контроля первичных и сводных данных системы административного мониторинга.....	32
1.3 Существующие инструментальные средства автоматизации при проведении административного мониторинга.....	39
1.3.1 Информационные системы управления.....	40
1.3.2 Системы аналитической обработки данных и прочие средства многомерного анализа.....	40
1.3.3 Системы интеллектуального анализа данных (data mining).....	40
1.3.4 Инструменты конечного пользователя для выполнения запросов и построения отчетов (query and reporting tools).....	41
1.4 Постановка задачи исследования.....	41
1.5 Выводы по первой главе.....	42
2 Формализация задач получения, анализа и контроля первичных и сводных данных при проведении административного мониторинга.....	44
2.1 Модель процесса переработки данных.....	44

2.2	Методическое обеспечение автоматизации процессов получения, анализа и контроля первичных и сводных данных	46
2.2.1	Методика настройки и получения сводных данных административного мониторинга	46
2.2.2	Методика построения процессов автоматизации анализа и контроля данных.....	48
2.2.3	Общая модель анализа и контроля данных	50
2.3	Моделирование структуры, правил формирования и содержимого выходных отчетных форм на основе шаблонов.....	51
2.3.1	Язык описания шаблонов выходных отчетных форм	52
2.3.2	Пример описания отчета с использование предложенного языка.....	54
2.3.3	Модель структуры отчета и правил его формирования	56
2.4	Моделирование критериев контроля и правил анализа данных	64
2.4.1	Формализация представления задач анализа и контроля данных в системе административного мониторинга.....	64
2.4.2	Модели правил анализа данных.....	70
2.4.3	Модели критериев контроля данных.....	74
2.5	Выводы по второй главе	78
3	Разработка и исследование алгоритмического и информационного обеспечения средств адаптации автоматизированной системы административного мониторинга	80
3.1	Разработка алгоритма формирования шаблона отчета	80
3.2	Методы анализа данных	83
3.3	Алгоритмы контроля данных.....	88
3.4	Автоматизация настройки задачи прогнозирования состояния объекта мониторинга.....	90
3.4.1	Задача построения классификационной функции для выбора метода прогнозирования	91

3.4.2	Исследование характеристик выходных наборов данных системы административного мониторинга.....	94
3.4.3	Методы прогнозирования временных рядов.....	98
3.4.4	Методы и алгоритмы нахождения характеристик выходных наборов данных системы административного мониторинга...	107
3.4.5	Нахождение оптимального набора характеристик временного ряда для автоматического выбора алгоритма прогнозирования.....	112
3.5	Выводы по третьей главе.....	118
4	Реализация и исследование прототипа средств адаптации автоматизированной системы административного мониторинга.....	120
4.1	Структурная схема подсистем генерации отчетов, анализа и контроля данных.....	120
4.2	Диаграммы состояний диалогового интерфейса пользователя.....	122
4.2.1	Интерфейс настройки подсистемы генерации отчетов.....	123
4.2.2	Интерфейс настройки правил анализа и критериев контроля данных.....	126
4.2.3	Интерфейс настройки задачи анализа и контроля данных.....	130
4.3	Архитектура программных средств подсистем генерации отчетов, анализа и контроля данных.....	132
4.4	Исследование прототипа и оценка эффективности функционирования средств адаптации подсистем генерации отчетов, анализа и контроля данных автоматизированной системы административного мониторинга.....	134
4.4.1	Оценка временных показателей по указанным интервалам....	138
4.4.2	Примеры.....	142
4.5	Выводы по четвертой главе.....	146
	ЗАКЛЮЧЕНИЕ.....	147
	Список литературы.....	149

ВВЕДЕНИЕ

В настоящее время производственные, экономические, социальные процессы все более приобретают распределенный характер. Создаваемые для реализации этих процессов организационные и организационно-технические системы (ОТС) имеют сложную структуру, их функционирование характеризуется огромным объемом показателей, динамически изменяющихся во времени. При принятии управленческих решений возникают проблемы, вызываемые, в том числе, сложностью своевременного получения полной информации о распределенном объекте управления. В этих условиях повышается актуальность применения распределенных автоматизированных систем мониторинга, обеспечивающих сбор, обработку и представление информации для оценки, контроля, прогнозирования или управления объектом мониторинга.

В ряду задач мониторинга можно выделить класс задач административного мониторинга, то есть мониторинга в организационных, организационно-технических и социальных системах, где основными источниками (операторами ввода) и потребителями (субъектами принятия решений) информации являются люди. В качестве источников данных также могут выступать существующие информационные системы предприятий или организаций.

Примерами задач административного мониторинга могут являться: мониторинг обеспеченности различных отраслей кадрами, мониторинг хода и результатов реализации целевых программ, мониторинг состояния основных фондов предприятия, мониторинг состояния жилищно-коммунальной сферы и др.

Несмотря на общую постановку задачи административного мониторинга, конкретные задачи достаточно специфичны. Специфика определяется как предметной областью, так и спецификой задач мониторинга в каждой конкретной области. В частности, специфика задачи

административного мониторинга определяется следующими основными факторами:

- разнообразие объектов учета;
- разнообразие исследуемых показателей (свойств) объектов учета;
- различные структуры связей объектов учета (иерархичность, отношения порядка и т.д.);
- разнообразие методик сбора данных и, как следствие, требуемых входных форм;
- разнообразие требований к обработке данных, структуре и содержанию выходных отчетов;
- разнообразие групп пользователей и их функций в процессе сбора и обработки данных.

Также нужно отметить, что даже в рамках одной конкретной задачи мониторинга (одной предметной области) зачастую требования к процедурам сбора, хранения и обработки данных могут меняться достаточно динамично:

- изменения способов, целей функционирования ОТС, их содержания, последовательности выполнения (например, изменение номенклатуры производимых изделий, переориентация производства);
- перемещения в пространстве отдельных элементов ОТС (например, оптимизация транспортной подсистемы корпорации, холдинга);
- перераспределения и децентрализации функций, задач, алгоритмов управления, информационных потоков между уровнями управления (модификация организационной структуры, внедрение АСУП);
- управления резервами (перераспределение кадровых, финансовых ресурсов, производственных мощностей);
- реконфигурации структур ОТС при ее деградации (например, задачи антикризисного управления).
- изменение требований законодательства, форм и состава необходимой отчетности, структуры объекта мониторинга и др.

В связи с этим система должна обладать свойством адаптации, т.е. оперативно реагировать на динамично меняющиеся требования. Поэтому становится очевидной потребность в адаптивных автоматизированных системах административного мониторинга (АСАМ).

Вопросы автоматизации процессов административного мониторинга и создание автоматизированных систем его поддержки рассмотрены в работах О.Б. Сладковой, Б.В. Соколова, Р.М. Юсупова, И.С. Константинова, А.В. Коськина, Д. Парк, С. Маккей, Е.Ф. Кодда, Д.А. Пospelова, А.Н. Фирстенко.

Однако вопросы построения и функционирования средств автоматизации в адаптивных системах административного мониторинга в настоящее время исследованы недостаточно полно, в частности отсутствуют унифицированные методы, модели и средства переработки данных в условиях изменяющихся требований к структуре и составу показателей, таким образом тема диссертационного исследования является актуальной.

Объект исследования – автоматизированные системы административного мониторинга.

Предмет исследования – модели и алгоритмы переработки данных, обеспечивающие адаптацию автоматизированной системы в условиях изменений требований к структуре и составу показателей административного мониторинга.

Цель исследования – повышение эффективности процесса адаптации автоматизированной системы административного мониторинга, при изменении требований к структуре и составу показателей.

Для достижения сформулированной цели были поставлены и решены следующие задачи:

1. Анализ особенностей проведения административного мониторинга, средств и методов его автоматизации.

2. Формализация задач получения, анализа и контроля первичных и сводных данных при проведении административного мониторинга с учетом требований адаптации.

3. Разработка и исследование алгоритмического и информационного обеспечения средств адаптации автоматизированной системы административного мониторинга.

4. Разработка прототипа и оценка эффективности функционирования средств адаптации подсистем генерации отчетов, анализа и контроля данных автоматизированной системы административного мониторинга.

Методы исследования. При решении указанных задач использовались: методы системного анализа, математического моделирования, дискретной математики, математической статистики, интеллектуальной обработки данных, оценки эффективности алгоритмов.

Достоверность и обоснованность научных положений, результатов, выводов и рекомендаций, приведенных в диссертационной работе, достигается за счёт: корректного применения известного математического аппарата; непротиворечивости и воспроизводимости результатов, полученных теоретическим и экспериментальным путём; соответствия результатов теоретических и экспериментальных исследований.

Научная новизна состоит в том, что:

1. Разработана модель процесса переработки данных, основанная на теории графов, отличающаяся возможностью многовариантной организации последовательности этапов процесса переработки данных.

2. С учетом модели процесса переработки данных разработаны модели: правил анализа данных, критериев контроля данных, структуры и содержания выходных отчетных форм; отличающиеся возможностью абстрагирования от конкретных объектов учета и показателей, и обеспечивающие процесс адаптации автоматизированной системы административного мониторинга.

3. Предложена методика построения процессов автоматизации анализа и контроля данных, базирующаяся на разработанных моделях и отличающаяся процедурой выбора источника данных.

4. Предложен алгоритм формирования шаблона отчета, отличающийся поуровневой настройкой объектов учета и основанной на модели структуры и содержания выходных отчетных форм.

5. Предложен алгоритм построения классификационной функции для автоматического выбора метода прогнозирования состояния объекта, обеспечивающий получение наиболее эффективного прогноза поведения объекта мониторинга с учетом его особенностей.

Практическая значимость заключается в разработке прототипа автоматизированной системы административного мониторинга, включающая в себя автоматизированные средства адаптации подсистем генерации отчетов, анализа и контроля данных системы административного мониторинга, который был применен в ФГБОУ ВПО Госуниверситет – УНПК (свидетельство о государственной регистрации № 2012660012 от 08.11.2012) .

Кроме того, полученные результаты диссертационного исследования внедрены в учебный процесс кафедры «Информационные системы» Госуниверситета-УНПК в рамках различных дисциплин, связанных с проектированием информационных систем. Разработан учебный курс «Проектирование распределенных адаптивных систем административного мониторинга» для направления подготовки магистров «Прикладная информатика»

Результаты диссертационной работы использованы при выполнении НИР (ГК № 02.740.11.0654 от 29.03.2010 г.) «Исследование и разработка теоретических основ построения и функционирования распределенных адаптивных систем административного мониторинга» ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009 – 2013 годы, гранта РФФИ 12-07-97528-р_центр_a «Интеллектуализация процессов

прогнозирования в сложных организационно-технических системах на основе данных административного мониторинга», гранта Президента Российской Федерации молодым ученым МК-3750.2012.9 «Научно-методические основы мониторинга процессов оказания электронных услуг населению». Разработанное программное обеспечение использовалось в составе программных информационно-аналитического и управляющего комплексов при выполнении НИР (ГК № П948 от 20.08.2009 г.) «Повышение эффективности управления интегрированными образовательными комплексами на основе информационно-аналитических ресурсов», НИР (ГК № 14.740.11.0591 от 05.10.2010 г. «Разработка универсальных инструментальных средств проектирования специализированных гибких модулей поддержки принятия решений в информационно-управляющих системах») той же ФЦП, применялось при выполнении работ по ГК № 598 от 12.09.08г. «Создание и внедрение системы оценки хода и результатов реализации федеральной целевой программы «Русский язык (2006-2010 годы)».

Апробация и публикации. Материалы диссертации докладывались на: Международной научно-технической Интернет-конференции «Информационные системы и технологии» (апрель-май 2011 г., г. Орел), V-ой Международной научно-технической конференции «Информационные технологии в науке, образовании и производстве» (май 2012 г., г. Орел), Международной молодежной конференции «Прикладная математика, управление и информатика» (2012 г., г. Белгород), Всероссийской молодежной конференции «Информационно-телекоммуникационные системы и технологии» (сентябрь 2012 г., г. Кемерово), Международной научно-технической Интернет-конференции «Информационные системы и технологии» (апрель-май 2013 г., г. Орел), VI-ой Международной научно-технической конференции «Информационные технологии в науке, образовании и производстве» (май 2014 г., г. Орел).

Положения, выносимые на защиту:

1. Модели: процесса переработки данных, правил анализа данных, критериев контроля данных, структуры и содержания выходных отчетных форм.

2. Методика построения процессов автоматизации анализа и контроля данных.

3. Алгоритмы формирования шаблона отчета, построения классификационной функции для выбора метода прогнозирования состояния объекта мониторинга.

4. Прототип подсистем генерации отчетов, анализа и контроля данных, обеспечивающих свойство адаптации автоматизированной системы административного мониторинга.

Публикации. По материалам диссертации имеется 11 публикации, в том числе: 5 статей в журналах из перечня ВАК, 6 публикаций в научных журналах и материалах конференций и 1 свидетельство о регистрации программ для ЭВМ.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения, списка литературы. Работа изложена на 162 страницах машинописного текста, включающего 24 рисунка, 5 таблиц, список литературы из 123 наименований.

Во введении дано обоснование актуальности темы диссертационного исследования, сформулированы цель и задачи работы, отмечена научная новизна и практическая значимость результатов, апробация работы и достоверность результатов, а также дано краткое изложение работы по главам.

В первой главе был проведен особенностей административного мониторинга средств и методов его организации, показано место административного мониторинга в системах управления, выявлены проблемы адаптации автоматизированной системы административного мониторинга и пути их решения. Были описаны задачи получения, анализа и

контроля первичных и сводных данных. Были рассмотрены существующие инструментальные средства автоматизации при проведении административного мониторинга.

Во второй главе представлена модель процесса переработки данных, методика настройки и получения сводных данных, методика построения процессов автоматизации анализа и контроля данных, язык описания шаблона выходных отчетных форм, модель структуры отчета и правил его формирования, модели правил анализа и критериев контроля данных.

В третьей главе представлены алгоритм формирования шаблона отчета, методы анализа данных, алгоритмы контроля данных и алгоритм построения классификационной функции для выбора метода прогнозирования.

В четвертой главе представлены результаты реализации и исследования прототипа подсистем генерации отчетов, анализа и контроля данных системы административного мониторинга. Разработаны и приведены диаграммы состояний диалоговых интерфейсов пользователя. В качестве основы архитектуры была выбрана модификация трехуровневой архитектуры, включающая компоненты хранилища, логики, представления и контроллер.

В заключении сформулированы основные выводы по результатам работы.

1 Анализ средств и методов организации административного мониторинга

1.1 Сущность административного мониторинга и его роль в принятии управленческих решений

1.1.1 Место автоматизированной системы административного мониторинга в системах управления

Согласно государственному стандарту автоматизированная система управления (АСУ) предназначена для обеспечения эффективного функционирования объекта управления путем автоматизированного выполнения функций управления и состоит из информационного, программного, технического, организационного, метрологического, правового и лингвистического обеспечения [1]. Функции АСУ в общем случае включают в себя следующие элементы (действия):

- планирование и (или) прогнозирование;
- учет, контроль, анализ;
- координацию и (или) регулирование.

Необходимый состав элементов выбирают в зависимости от вида конкретной АСУ [1].

По выполняемым функциям и виду управляемого процесса можно выделить АСУ: административно-организационные, технологические (для управления технологическими процессами) и интегрированные [1].

Среди многообразия современных автоматизированных систем управления производством (АСУП) присутствуют такие, обязательной функцией которых является ручная переработка данных. При этом, учитывая распределенность рассматриваемых объектов управления, существенно повышается значимость эффективной реализации обеспечивающих функций АСУП – получения сводных данных в нужном разрезе и уровне абстракции, анализа и контроля данных.

Таким образом, область наших интересов лежит в части АСУП и их подсистем, обеспечивающих следующие функции:

- сбор и контроль полученных данных;
- приведение к нужному формату, первичную обработку данных;
- организацию хранения и предоставления доступа к сохраненным данным;
- анализ данных мониторинга;
- генерацию запросов;
- генерацию отчетных форм, визуализацию отчетов.

Чтобы показать место системы мониторинга, реализующей функции сбора, хранения и переработки данных, в АСУ, напомним классический контур управления АСУ [2]. Поступающее на объект управляющее воздействие изменяет его состояние. Полученное состояние объекта управления регистрируется измерительным механизмом, который оценивает его и передает субъекту управления. Субъект управления после анализа полученной оценки состояния выдает вектор управления, который принимает и исполняет регулятор. Исполнение заключается в направлении на объект управления управляющего воздействия (Рисунок 1.1).



Рисунок 1.1 – Контур управления в АСУ

В отличие от автоматических систем, в которых после их запуска роль человека сводится к контролю за работой системы, в автоматизированных системах человек является главным определяющим звеном этих систем, поэтому при проектировании АСУ необходимо учитывать такие «человеческие» факторы, как пропускная способность человека, скорость реакции, допустимые объемы перерабатываемой информации. В этой связи даже наличие полной (с требуемой долей абстракции) информации об объекте управления (мониторинга) не гарантирует возможности принятия качественного управленческого решения. Это особенно актуально для сложных организационно-технических систем, когда само понятие «сложность» определяется как невозможность целостного одновременного их охвата некоторым наблюдателем (Уильям Росс Эшби).

Поэтому процесс принятия решений в той или иной мере должен сопровождаться и подкрепляться использованием соответствующих формальных моделей. Именно процессы моделирования и прогнозирования позволяют из огромных массивов исходных данных (информационных ресурсов) получать информационно-аналитические ресурсы [3], отличающиеся более высоким уровнем абстракции и пригодные для анализа, оценки и принятия управленческих решений.

В работах Коськина А.В., посвященных управлению в сложных ОТС, выделяются следующие основные этапы формирования управления.

- 1 Возникновение проблемы.
- 2 Постановка задачи.
- 3 Исследование внутренней структуры ОТС.
- 4 Исследование состояния внешней среды.
- 5 Комплексное применение ряда методов научных исследований для получения множества предварительных вариантов управленческих решений.
- 6 Выбор окончательного решения.

Очевидной, что третий и четвертый этапы с точки зрения общей методики организации процесса административного мониторинга [4],

соответствуют процедурам сбора и организации хранения информации в системе административного мониторинга.

Пятый этап предполагает генерацию альтернатив управляющих воздействий. Именно в рамках этого этапа происходит анализ состояния объекта управления и возникшей проблемной ситуации. Лицо, принимающее решение, на основе информации, полученной при сборе данных должно выявить причину проблемы и только после этого формулировать цель управления и переходить к генерации и оценке альтернатив. Вопрос автоматизации оценки альтернатив выходит за рамки исследования и лежит в плоскости, скорее всего, имитационного моделирования. Однако вопросы анализа состояния объекта управления, то есть вопросы аналитической обработки данных мониторинга могут быть решены на данном этапе системой административного мониторинга за счет реализации подсистемы анализа и контроля данных.

Задачи контроля данных должны решаться в текущем режиме. Их цель – предупредить появление проблемной ситуации. Система административного мониторинга является источником больших объемов данных о состоянии системы, поступающих на периодической основе. Их текущий оперативный анализ лицом, принимающим решения, невозможен. Однако возможно заранее определить ряд показателей (первичных или агрегированных), которые будут служить индикаторами изменений, происходящих в системе. На значения этих показателей могут быть наложены ограничения, соответствующие «нормальному» режиму функционирования системы. Причем, логическая сложность этих ограничений невысока, так как их цель – лишь информировать о выходе показателей за нормативные значения или о приближении к их границам.

Место АСАМ в системах управления[5] показано на рисунке 1.2, где представлены два контура управления: контур управления объектом (основной контур) и контур управления системой административного мониторинга.



Рисунок 1.2 – Место АСАМ в системах управления

Основной контур включает управляющее воздействие, направленное на объект управления (мониторинга), которое оцениваются по его данным, поступающим на вход системы административного мониторинга.

Результатом функционирования системы административного мониторинга является формирование информации о состоянии объекта управления в виде отчетов, как в текущее время, так и в ретроспективе. Также система административного мониторинга преобразует информацию о состоянии об объекте управления в информационно-аналитические ресурсы, отличающихся более высоким уровнем абстракции и пригодными для анализа, оценки и принятия управленческих решений. Это происходит за счет имеющейся подсистемы анализа и контроля данных в АСАМ.

На основании информации о состоянии объекта управления или информационно-аналитических ресурсов, характеризующих состояние объекта управления, в системе выработки управляющих воздействий, включающей обычно лицо, принимающее решения, и систему поддержки принятия решений, формируется вектор управляющей информации, такой, что ее доведение до элементов ОТС приведет систему в целевое состояние.

Роль исполнительного механизма, оказывающего управляющее воздействие на объект управления, выполняет система реализации управляющих воздействий, то есть формирования, согласования и контроля исполнения организационно-распорядительных документов. В случае автоматизированной реализации данного процесса такую систему называют обычно системой электронного документооборота.

Контур управления (адаптации) системы административного мониторинга используется при изменении внутренних и внешних условий функционирования организационно-технической системы, при решении стратегических задач управления, связанных с изменением организационной структуры, перераспределения ресурсов, реорганизации производственного процесса и т.п. При решении таких задач зачастую данных об объекте управления недостаточно для принятия решения. В этом случае на основе анализа текущего состояния объекта управления, которое характеризует не только деятельность организационно-технической системы, но и функции системы административного мониторинга формируется вектор управляющей информации, преобразование которого в вектор управления системой мониторинга приводит к изменению ее функций сбора данных и функций хранения и переработки данных (изменяется состояние объекта управления и ИАР, характеризующих состояние объекта управления). Необходимо отметить, что изменение состояния объекта управления и ИАР, характеризующих состояние объекта управления, может осуществляться и при неизменных входных информационных потоках системы мониторинга. Такое изменение будет соответствовать адаптации процедур получения сводных данных, анализа и контроля данных в системе мониторинга.

1.1.2 Основные понятия и положения организации мониторинга

1.1.2.1 Мониторинг

Мониторинг – специально организованное, систематическое наблюдение за состоянием объектов, явлений, процессов с целью их оценки, контроля, прогноза и управления [6].

Мониторинг является информационной базой для выполнения функций управления, но и сам реализуется при помощи общих функций управления [7].

С точки зрения технических наук под *системой мониторинга* понимается взаимосвязанная совокупность элементов, обеспечивающая осуществление мониторинга [7].

Можно выделить следующие составляющие структуры системы мониторинга:

- субъекты мониторинга;
- объект мониторинга;
- комплекс показателей и индикаторов мониторинга;
- инструментарий мониторинга;
- процедура мониторинга.

1.1.2.2 Субъект мониторинга

Субъектами мониторинга являются носители функций систематического наблюдения за состоянием объекта мониторинга, то есть непосредственно функций мониторинга. Условно их можно разделить на две группы [8]:

- производящие сбор информации;
- выполняющие обработку информации.

1.1.2.3 Объект административного мониторинга

Объектом мониторинга в общем случае является некоторый объект или процесс реального мира. Виды объектов мониторинга разнообразны, их

масштабность, динамичность, сложность существенно отличается, что отчасти и обуславливает разнообразие видов и систем мониторинга.

«Следует разграничить объект административного мониторинга и объект управления АСУ. Объект мониторинга уже, чем объект управления интегрированной АСУ, функциональные подсистемы которой могут охватывать все сферы основной и вспомогательной деятельности ОТС (в том числе и организационное управление).» [7].

Объектами мониторинга в общем случае могут являться как сложные системные объекты (например, отрасль промышленности, производственное объединение, корпорация, холдинг и их деятельность, здоровье персонала, экологическое состояние территории предприятия и пр.), так и достаточно локальными (например, процесс внедрения в производство нового изделия, внедрение энергосберегающей технологии на предприятии и пр.) [9].

С точки зрения административного мониторинга на предприятии объектом мониторинга могут являться процессы (основные, вспомогательные, обеспечивающие) или структурные элементы предприятия (средства производства, объекты инфраструктуры, элементы организационной структуры и др.).

Однако, несмотря на разнообразие объектов мониторинга, можно выделить общие свойства, которые объединяют все эти разнородные объекты, принадлежащие даже к различным сферам деятельности [7] [9]:

- объекты мониторинга динамичны, находятся в постоянном развитии, они подвержены влиянию внешних воздействий;
- реализация мониторинга предполагает организацию слежения за объектом;
- организация слежения предусматривает отбор обоснованных показателей и индикаторов, характеризующих объект мониторинга;
- слежение осуществляется путем непосредственного или опосредованного измерения параметров объекта.

1.1.2.4 Комплекс показателей и индикаторов мониторинга

Осуществление мониторинга невозможно без четкого определения показателей и индикаторов исследуемого объекта мониторинга. Понятие показателя используется достаточно широко в различных областях. В общем случае показатель выступает методологическим инструментом, обеспечивающим возможность проверки теоретических положений с помощью эмпирических данных [10]. В рамках рассматриваемой тематики под *показателем* понимается обобщенная характеристика свойства объекта мониторинга. Обычно выделяют следующие виды показателей:

- качественные показатели, фиксирующие наличие или отсутствие определенного свойства;
- количественные показатели, фиксирующие меру выраженности, развития определенного свойства.

Индикатор – доступная наблюдению и измерению характеристика изучаемого объекта. Индикаторами замещают, с помощью индикаторов обнаруживают и представляют другие характеристики (показатели) исследуемого объекта, обычно недоступные наблюдению или неудобные (непригодные) для интерпретации. Таким образом, можно говорить, что значения индикаторов являются результатом изучения, обобщения и сравнения значений некоторых показателей [11].

Комплекс *показателей и индикаторов* обеспечивают представление о состоянии системы, о качественных и количественных изменениях в ней, необходимое и достаточное для реализации функций управления.

1.1.2.5 Процедура мониторинга

Процедура мониторинга представляет собой совокупность операций, проводимых в ходе мониторинга. Например, подготовка мониторинга, сбор, обработка, анализ и представление информации, обеспечение мониторинговых процедур [12].

Основная сфера практического применения мониторинга – это управление, а, точнее, информационное обслуживание управления в различных областях деятельности [12].

1.1.3 Существующие виды систем мониторинга

Разнообразие областей применения и существование большого количества различных систем мониторинга обуславливает необходимость их определенного упорядочения. С этой целью задачи мониторинга и системы мониторинга можно классифицировать по нескольким критериям. К таким критериям можно отнести [13]:

- назначение или область применения мониторинга;
- средства, используемые для сбора исходной информации;
- ориентация на конкретного пользователя;
- направленность на реализацию задач функционирования или задач развития;
- виды управленческих задач;
- и другие.

1.1.3.1 Классификация по назначению или области применения мониторинга

В зависимости от области применения результатов мониторинга можно выделить радиолокационный, авиационный, космический, спутниковый, инструментальный, педагогический, психологический, социологический, медицинский, статистический, административный [14] [15] [16] [17] и другие виды мониторинга.

1.1.3.2 Классификация на основании способа сбора информации

На основании способа сбора информации можно выделить три группы видов мониторинга.

В первую группу входят те виды мониторинга, в процессе осуществления которых возможно непосредственное описание объекта мониторинга без каких-либо измерений, с использованием технологии структуризации результатов, построения схемы и технологии сбора информации (например, мониторинг средств массовой информации, текущего законодательства, выборов).

Вторую группу составляют виды мониторинга, в процессе которых осуществляется непосредственное физическое измерение параметров объекта (например, мониторинг шума, уровня моря, коррозии металлов, компьютерных сетей).

Третья группа включает виды мониторинга, в ходе которых измерение параметров объекта проводится опосредованно с использованием системы хорошо разработанных и общепринятых критериев или индикаторов (например, мониторинг качества воздуха, воды, сердечной деятельности, почвенно-химический мониторинг).

Четвертую группу составляют те виды мониторинга, при проведении которых нет возможности использовать существующие системы критериев и индикаторов. В этом случае необходимо опосредованное изменение и привлечение технологий научного исследования, обеспечивающих преобразование результатов непосредственных измерений в требуемый вид (например, мониторинг санитарно-гигиенический, социально-политический, социально-экономический) [18].

1.1.3.3 Классификация видов мониторинга в соответствии с ориентацией результата на конкретного пользователя

Можно выделить три группы, отличные по количеству пользователей и интенсивности использования результатов мониторинга соответствующим пользователем.

Первую группу составляют *виды мониторинга, ориентированные на общество в целом*. Их целью может являться, например, формирование

общественного мнения. Виды мониторинга первой группы немногочисленны. Ознакомление пользователя с результатами мониторинга в этом случае осуществляется через средства массовой информации, в том числе и электронные.

Вторая группа включает *виды мониторинга, ориентированные на специалистов соответствующих областей* деятельности. Это наиболее многочисленная группа. К ней принадлежит большинство существующих систем мониторинга. При этом сами группы специалистов, для которых предназначены результаты каждого конкретного мониторинга, могут быть как достаточно малы, так и очень многочисленны. Основными способами распространения получаемой в ходе такого вида мониторинга информации являются специализированные издания, в том числе периодические, интернет-ресурсы, подписка.

Третья группа включает *виды мониторинга, пользователями которых являются конкретные органы управления, руководители, отдельные структуры*. Средством распространения информации, получаемой в ходе такого рода мониторинга, являются аналитические отчеты, рекомендации, проекты, которые обычно не имеют широкого распространения [19].

1.1.3.4 Классификация по видам управленческих задач

1. Мониторинг, направленный на задачу *функционирования* – мониторинг локального характера, посвященный одной задаче или одной проблеме. Реализация этого мониторинга сопровождает существование объекта мониторинга на протяжении реализации им определенной функции, в том числе неограниченное время;

2. Мониторинг, направленный на задачу *развития*. Текущие задачи развития и, соответственно, предмет изучения этого мониторинга существует некоторое время. После того как задача развития решена, мониторинг прекращается [19].

1.1.3.5 Классификация по характеру использования мониторинговых данных

Информационно-поисковые системы реализуют функции ввода, систематизации, хранения и выдачи информации по запросу пользователя без сложных преобразований данных.

Информационно-решающие системы осуществляют все операции переработки информации по определенному алгоритму. Среди них можно провести классификацию по степени воздействия выработанной результатной информации на процесс принятия решений: управляющие и советующие [20].

1.1.3.6 Выводы по результатам классификации

Проанализировав существующие виды мониторинга можно охарактеризовать предлагаемые средства организации мониторинга с точки зрения приведенных выше критериев классификации мониторинга.

Система мониторинга в состоянии обеспечивать мониторинг направленного на задачи функционирования и развития. Однако если учитывать затрагиваемые в данном исследовании вопросы адаптации систем мониторинга, то более уместно говорить о мониторинге, направленном на задачу развития. Это обуславливается тем, что применение более сложных адаптивных систем административного мониторинга [7] актуально в случае решения периодически возникающих краткосрочных и среднесрочных задач сбора и переработки данных с меняющимся составом и структурой показателей.

1.1.4 Специфика административного мониторинга

Основываясь на изложенном выше, опишем специфические особенности административного мониторинга. Административный мониторинг рассматривается как неотъемлемая составляющая процесса

управления крупными производственными объектами и сложными организационными процессами на производстве. При этом целью производственного управления является такое воздействие на организационную и техническую системы предприятия, которое приведет их в требуемое состояние. То есть объектом управления является организационно-техническая система предприятия. Исполнительным механизмом в таком случае являются люди, обеспечивающие производственные процессы на предприятии [7].

Объект управления и объект административного мониторинга в данном случае, как и субъект управления, и субъект административного мониторинга, могут быть нетождественны. Как в случае административного управления, так и в случае административного мониторинга субъектами являются лица, принимающие решения. В ряде случаев лицо, принимающее решение (ЛПР) в процессе управления, формулирует цель, задачи и другие требования к процедуре мониторинга, и, соответственно, является субъектом мониторинга. Однако более вероятной является ситуация, когда субъект управления формулирует только цель и задачи мониторинга, а принятие решений и управление процедурой мониторинга, другое лицо – субъект мониторинга [7].

Целью административного мониторинга является получение объективной, полной, достоверной и актуальной информации об объекте мониторинга, необходимой для оценки, прогнозирования состояния объекта и формирование вектора управления. Для достижения заявленной цели необходимо, но недостаточно решить задачу сбора данных и переработки данных. Административный мониторинг является более сложным процессом, реализация которого подразумевает решение нескольких задач [4].

1. Подготовка процедуры мониторинга. Данную задачу можно разделить на две подзадачи: организационную и техническую подготовку. Организационная подготовка заключается в разработке или модификации нормативно-распорядительной документации. Техническая составляющая

заключается в подготовке программно-аппаратных средств системы административного мониторинга.

2. Сбор данных. В связи со спецификой административного мониторинга сбор данных заключается во вводе операторами соответствующей информации посредством диалоговых интерфейсов. Также возможна ситуация, когда данные импортируются из подсистем АСУП.

3. Хранение данных. В процессе принятия управленческих решений зачастую для формирования обоснованного вектора управления необходимо иметь не один временной срез информации о состоянии объекта управления, а данные за определенный диапазон периода наблюдения.

4. Переработка данных. Во-первых, необходима первичная обработка данных, Во-вторых, собранная и прошедшая первичную обработку информация может не обладать свойством полезности, так как не может быть одновременно охвачена ЛПР. Вторым этапом является переработка данных, приводящая информацию об объекте мониторинга в вид, пригодный для оценки, прогнозирования и принятия управленческих решений.

На основе изложенного выше можно сформулировать определение административного мониторинга. Административным мониторингом называется вид мониторинга, предполагающий оперативный сбор информации о состоянии объекта управления из различных источников с последующей ее переработкой до необходимого уровня абстракции и дальнейшего предоставления ее в качестве исходной для различных систем поддержки принятия решений и/или лицу принимающему решение в контурах управления ОТС, в условиях изменяющихся требований к структуре и составу показателей.

1.2 Проблемы процесса адаптации и пути их решения

1.2.1 Отдельные проблемы применения больших сложных АСУ

Проектирование и построение систем автоматизации достаточно трудоемкий и длительный процесс. Большинство АСУП предметно- или

проблемно-ориентированы [21], то есть созданы для решения конкретной производственной или организационной задачи, либо для автоматизации общих процедур, происходящих в определенной предметной области. В архитектурном смысле подобного рода АСУП представляют собой отображение процессов предметной области объекта автоматизации. Делается это путем определения модели предметной области «на языке реализации автоматизированной системы» [22]. Предметная область описывается достаточно однозначно, и каждому объекту реальности жестко соответствует абстрактный объект, создаваемый при проектировании.

Работу проблемно-ориентированных АСУП можно рассматривать как поиск выхода из сложной ситуации, они ориентированы на отображение определенного среза предприятия. Следовательно, подсистемы сбора, хранения и обработки данных проектируются относительно требований к необходимой отчетной информации [23].

Удобство предметно-ориентированного подхода ограничивается этапом проектирования, когда разработчики и специалисты рассматриваемой области используют один язык для описания своих требований и целей, ориентированный на предметную область. Так же предметно-ориентированную архитектуру легко тестировать.

Однако, несмотря на технические преимущества, данный подход не позволяет оперативно модифицировать систему мониторинга в дальнейшем и ее ориентировать на другие задачи управления из-за временных затрат реструктуризация базы данных, модификации и/или написания запросов, модулей, исходного кода.

1.2.2 Организация процесса адаптации системы административного мониторинга

Если рассматривать классическое значение термина «адаптивность», принятое в теории автоматического управления и кибернетике, то речь идет о

свойстве системы управления автоматически изменять свои параметры (функцию управления) или свою структуру. Соответственно выделяют самонастраивающиеся и самоорганизующиеся системы. В приведенных выше способах управления адаптациями термин «адаптивность» полностью соответствует автоматической адаптации.

Термин же «адаптируемость» подразумевает изменение параметров, функций или структуры системы с помощью внешних воздействий. Поэтому, если рассматривать систему административного мониторинга узко – как программно-технический комплекс, то ее можно назвать адаптируемой. Причем в классической трактовке эксплуатационное требование адаптируемости программного обеспечения означает свойство программного продукта, обеспечивающее простоту его модернизации. В настоящее время в программной инженерии адаптируемость все чаще понимается как возможность быстрой перенастройки программного обеспечения без изменения текста программ и физической структуры данных.

Если же, основываясь на невозможности функционирования системы административного мониторинга без участия человека, рассматривать ее в более широком смысле и считать человеко-машинной системой, то она обладает свойством адаптивности в значении, близком к приведенному выше.

Возвращаясь к системе мониторинга (СМ) можно сказать, что даже в рамках одной предметной области может наблюдаться существенное разнообразие объектов мониторинга. Таким образом, требование адаптивности следует соотноситься объектом мониторинга. Это приводит к необходимости формулировки некоторого абстрактного, формализованного представления объекта мониторинга в виде множества абстракций объектов и процессов реального мира (будем называть их объектами учета) и информационных связей как внутренних, так и с внешней средой.

Физически специфика представления объекта мониторинга и его связей будет обеспечиваться подсистемой хранения структуры объектов учета.

Именно подсистема хранения обеспечивает адаптируемость СМ, учитывая все многообразие объектов учета, связей между ними (иерархических, сетевых), множества исследуемых показателей объектов учета.

В качестве способа управления адаптацией в СМ используется структурная адаптация. На специализированном интерфейсе производится настройка объекта мониторинга и его связей. При этом алгоритмы работы с абстрактным объектом мониторинга остаются неизменными.

По уровню адаптация происходит на уровне метаданных, однако модель структуры хранения должна быть организована таким образом, что при проведении адаптации меняются не физические метаданные, а метаданные более абстрактного уровня, реализованные через физические «линейные» сущности.

В диссертационной работе Кравцовой Н.А. «Автоматизация процессов хранения и сбора данных при проведении административного мониторинга» [24] решена задача автоматизации процессов хранения и сбора данных: предложена модель структуры хранения данных административного мониторинга, обеспечивающая инвариантное представление объекта мониторинга и гибкую настройку при изменении требований к процедуре мониторинга; разработаны модель процесса синтеза конкретной модели хранения данных административного мониторинга, модель процесса регистрации значений показателей и реализующие данные процессы алгоритмы; разработаны модели диалоговых интерфейсов чтения и записи данных, а также алгоритм отображения модели хранения на модель интерфейса, обеспечивающий возможность автоматического синтеза (генерации исходного текста) диалоговых интерфейсов ввода данных на основании модели хранения. Однако, вопросы автоматизации получения, анализа и контроля первичных и сводных данных административного мониторинга не вошли в рамки данной диссертационной работы.

Для обеспечения оперативной адаптации автоматизированной системы административного мониторинга задачи получения, анализа и контроля

первичных и сводных данных должны быть формализованы в терминах специальной модели хранения данных, основные элементы которой приведены ниже.

Информация, хранимая в системе мониторинга уже определенным образом структурирована [25] [26].

Основными информационными объектами абстракциями модели структуры хранения являются следующие множества: V_t – множество типов, P – множество показателей, V_o – множество экземпляров объектов учета.

Элементы приведенных выше множеств вступают в отношения ассоциации и агрегации. Описание соответствующих множеств приведено ниже [25].

Множество родительских связей типов – подмножество декартова произведения $V_t \times V_t$. $E_t = \{(t_i, t_j) \mid (t_i \hat{I} V_t) \& (t_j \hat{I} V_t)\}$, где $e_t = (t_i, t_j)$ – элемент множества E_t , упорядоченная пара типов «родитель (i) – потомок (j)» соответственно.

$E_{to} = \{(t, o) \mid t \hat{I} V_t \& o \hat{I} V_o\}$ – множество отношений ассоциации экземпляров объекта учета к типам; e_{to} – отношение ассоциации объекта учета o с типом t .

Множество всевозможных наборов показателей – $2^P = \{X_P \mid X_P \hat{I} P\}$.

R_{tp} – подмножество декартова произведения множеств ($2^P \times V_t$) – подмножество показателей, ассоциированных с некоторым типом объектов учета: $R_{tp} = \{(X_P, t) \mid X_P \hat{I} P \& t \hat{I} V_t\}$.

Элементы множества E_{to} – типизированные объекты учета – попарно связываются между собой связью «потомок-родитель», создавая упорядоченные пары элементов множества E_{to} . Возможность существования такого отношения определяется связями соответствующих типов, таким образом получаем множество связей экземпляров объектов учета:

$E_o = \{(e_{toi}, e_{toj}) \mid e_t = (t_i, t_j) \& e_{toi} = (o_i, t_i) \& e_{toj} = (o_j, t_j)\}$.

Отношение типизированных ОУ к наборам показателей:

$$\mathbf{R}_{po} = \mathbf{E}_{to} \circ \mathbf{R}_{tp} = \{(\mathbf{o}, \mathbf{X}_P) \mid \mathbf{o} \in \mathbf{V}_o \ \& \ \mathbf{X}_P \subset \mathbf{P} \ \& \ \exists \mathbf{t} \in \mathbf{V}_t, \mathbf{e}_{to} = (\mathbf{t}, \mathbf{o}) \ \& \ \mathbf{r}_{tp} = (\mathbf{X}_P, \mathbf{t})\}.$$

1.2.3 Задачи получения, анализа и контроля первичных и сводных данных системы административного мониторинга

Основной функцией системы мониторинга с точки зрения представления результатов является первичная обработка на основе заданных правил и генерация простых отчетов по результатам (получение сводных данных) мониторинга. Под простыми понимаются отчеты, представляющие собранную информацию об объектах учета и значениях показателей за указанный период в определенном разрезе с возможностью простейшей статистической обработки (определение средних, максимальных и минимальных значений показателей). Кроме того, на систему мониторинга могут возлагаться функции анализа и параметрического контроля собранных данных. В данном случае под анализом данных понимается агрегирование данных мониторинга по определенным правилам, нахождение статистических оценок и другие виды обработки, повышающие уровень абстракции результатов мониторинга до приемлемого для единовременного охвата и принятия управленческих решений. Под контролем данных понимается их оценка на соответствие некоторым показателям, заданным лицом, принимающим решения. Контролю могут подвергаться как детальные данные мониторинга, так и агрегированные. При этом должна обеспечиваться возможность использовать как четкие критерии соответствия (или правила логического вывода), так и нечеткие.

Задачи получения, анализа и контроля первичных и сводных данных системы административного мониторинга соответствуют технологическому процессу переработки данных. По ГОСТ Р 51170-98 [27] технологический процесс переработки данных – совокупность действий направленных на изменение состояния данных.

1.2.3.1 Получение сводных данных

Задача получения сводных данных в виде отчетов соответствует технологическим операциям обработке данных и обобщению данных. По ГОСТ Р 51170-98 обработка данных – технологическая операция в результате которой изменяет свое значение хотя бы один из показателей, характеризующих состояние данных (объем данных при этом не меняется). По ГОСТ Р 51170-98 обобщение данных – технологическая операция, в результате которой из исходных данных получают новые данные уменьшенного объема.

В любой системе технология реализации подсистемы генерации отчетов существенным образом определяется применяемой технологией хранения данных. В случае адаптивной системы административного мониторинга используется особая модель хранения данных на физическом уровне, обеспечивающая инвариантность представления данных логического уровня. Поэтому ни один из существующих генераторов отчетов и ни одна из технологий генерации не могут быть напрямую перенесены в рассматриваемую систему. Однако, базовые принципы на которых основывается генерация отчета в системах различных классов могут быть (с определенной модификацией) использованы при построении подсистемы генерации отчетов адаптивной системы административного мониторинга.

Сначала рассмотрим понятие отчета. Отчет в некоторой системе управления – это документ. В [28] документ определяется как материальный объект, содержащий информацию в зафиксированном виде и специально предназначенный для её передачи во времени и пространстве.носителем информации может быть как бумага, так и различные виды запоминающих устройств. Большая часть отчетов (если рассматривать задачу вывода данных мониторинга) относится к информационно-справочным документам. Однако, если рассматривать процесс использования отчета в целом, то помимо информационной функции он выполняет и функцию обеспечения процесса управления.

Отчет в общем случае представляет собой сложную информационную структуру, характеризующуюся множеством различных свойств (реквизиты, формат, совокупность полей, единицы измерения, правила расположения данных и т.д.). При этом, учитывая разнообразие объектов учета системы мониторинга и процессов управления в организационно-технических системах различных классов, множество отчетов будет характеризоваться существенным разнообразием. Однако, несмотря на это в общем виде любой отчет определяется двумя составляющими – формой и содержанием. Содержание отчета в системе административного мониторинга – есть отображение логической модели хранения данных, отражающее семантический аспект информации без конкретных требований к ее представлению на физическом носителе. Содержание отчетов определяется существующей или формируемой системой управления ОТС, в ходе автоматизации процессов управления. Именно системой управления устанавливаются требования к составу и содержанию отчетных форм, способам их применения в управленческой деятельности. Задача определения формы отчета в некотором смысле является технической. Однако, недооценка важности формы отчета достаточно сильно может отразиться на скорости реализации процессов управления, особенно в сложных системах, характеризующихся большим количеством разнотипных показателей.

На основе изложенного выше можно говорить, что процесс генерации отчета в системе административного мониторинга – это преобразование данных об объекте мониторинга в информацию о нем, представляемую в структурированном виде.

Теоретически задача генерации отчета включает в себя три подзадачи:

- выборка данных в соответствии с определенными множествами исходных данных и условиями выборки;
- обработка данных, то есть формирование представления об объекте на логическом уровне;

– визуализация информации отчета, то есть его оформление и вывод пользователю.

Во всех развитых универсальных генераторах отчетов в то или ином виде для создания отчета используются шаблоны. Это позволяет использовать единожды определенную структуру и/или правила оформления для периодической генерации однотипных отчетов. Зачастую шаблон определяет как содержание отчета, так и его форму, находящиеся в неразрывной связи. Однако для сложных структур данных, которые могут быть отображены в различном виде шаблон содержимого может быть отделен от шаблона формы отчета.

Для задания шаблона отчета в большинстве существующих генераторов отчетов используются специализированные или универсальные языки, позволяющие определять иерархические структуры (основываясь на возможности представления любой табличной формы в виде дерева, например, см. структурные диаграммы Джексона [29]). Причем язык может рассматриваться как инструмент хранения и интерпретации информации о шаблоне, так и как инструмент описания шаблона пользователем. Первый вариант более предпочтителен вследствие снижения требований к квалификации пользователя, определяющего структуру шаблона, однако, более сложен в реализации. Поэтому в большинстве случаев (когда шаблоны отчетов изменяются редко) реализуется первый вариант. Однако, в любом случае при использовании шаблонов в задаче генерации отчета появляется еще одна подзадача – создание и редактирование шаблонов отчетов.

По результатам проведенного анализа можно сформулировать базовые принципы построения подсистемы генерации отчетов адаптивной системы административного мониторинга.

- 1 Процесс генерации отчета должен включать следующие этапы:
 - создание или выбор готового шаблона отчета;
 - задание модели структуры отчета и правил его формирования;

– выборка данных из базы данных системы административного мониторинга и их обработка;

– построение и вывод отчетной формы.

2 Шаблон отчета должен определять его содержание, то есть типы объектов учета или конкретные объекты учета, их связи в иерархии и выводимые для них показатели. Такая трактовка шаблона обусловлена возможной сложностью информационной модели объекта учета и возможностью осуществлять отображение одного и того же информационного массива различными способами.

3 Для представления шаблона отчета должен использоваться специализированный язык, позволяющий описывать иерархические структуры данных. Язык является как средством хранения шаблона и его интерпретации (то есть предполагается наличие визуального диалогового интерфейса создания и редактирования шаблона), так и, возможно, средством описания шаблона пользователем.

4 Форма отчета, то есть структура разметки таблицы и правила отображения данных, задается моделью структуры отчета и правил его формирования. Для краткости в дальнейшем будем называть данную модель стилем визуализации.

5 С целью упрощения типовых задач формирования отчетных форм в системе должны быть предусмотрены типовые шаблоны, определяющие абстрактные типы объектов учета и их показатели, и сопоставленные с ними стили визуализации. В этом случае для создания отчета пользователю необходимо конкретизировать лишь множество исходных данных.

6 Выборка данных осуществляется на основе выбранного пользователем шаблона отчета. В ходе выборки должны использоваться реализованные в блоке управления данными функции чтения и записи. По результатам выборки в памяти строится древовидная структура данных, содержащая всю исходную информацию, необходимую для построения

отчета. При обработке данных в структуру добавляются результаты обработки.

7 Непосредственная генерация отчетной формы осуществляется на основании полученной структуры данных и заданного стиля визуализации.

8 Должны быть реализованы табличная и графическая форма представления отчетов.

9 Подготовленные отчеты должны сохраняться в постоянной памяти с целью минимизации использования вычислительного ресурса сервера при повторной генерации отчета.

1.2.3.2 Анализ данных

Задача анализа данных соответствует интеллектуальному анализу данных, в англоязычной литературе data mining.

Data Mining – одной из современных аналитических метатехнологий, предназначенной для переработки сырой информации с целью получения продуктивных знаний. Data mining – это процесс обнаружения в сырых данных ранее неизвестных нетривиальных практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности [30].

В настоящее время широко применяются методы интеллектуальной обработки данных, объединенные названием Data Mining [31]. Термин Data Mining часто переводится как добыча данных, извлечение информации, интеллектуальный анализ данных, средства поиска закономерностей, анализ шаблонов, раскопка знаний в базах данных.

Методы Data Mining помогают решить многие задачи, с которыми сталкивается аналитик. Из них основными являются: классификация, регрессия, кластеризация, поиск ассоциативных правил и прогнозирование [31].

Классификация – наиболее простая и распространенная задача Data Mining. В результате решения задачи классификации обнаруживаются

признаки, которые характеризуют группы объектов исследуемого набора данных – классы; по этим признакам новый объект можно отнести к тому или иному классу.

Задача регрессии подобно задаче классификации позволяет определить по известным характеристикам объекта значение некоторого его параметра. В отличие от задачи классификации значением параметра является не конечное множество классов, а множество действительных чисел.

Задача кластеризации заключается в поиске независимых групп (кластеров) и их характеристик во всем множестве анализируемых данных. Решение этой задачи помогает лучше понять данные. Кроме того, группировка однородных объектов позволяет сократить количество необходимых итераций процедуры анализа.

При поиске ассоциативных правил целью является нахождение чистых зависимостей (или ассоциаций) между объектами и событиями. Найденные зависимости представляются в виде правил и могут быть использованы как для лучшего понимания природы анализируемых объектов, так и для предсказаний появления событий.

Задача прогнозирования, может считаться одной из наиболее сложных. Она требует тщательного исследования исходного набора данных и методов, подходящих для анализа. Прогнозирование, в широком понимании, определяется как опережающее отражение будущего, а его целью является предсказание будущих событий.

1.2.3.3 Контроль данных

Задача контроля данных АСАМ соответствует технологической операции контроля данных. По ГОСТ Р 51170-98 контроль данных – технологическая операция, состоящая в сравнении значений показателей, характеризующих состояние данных с определенными значениями. Такими значениями могут являться [32]:

- наименьшее и наибольшее значения показателя (минимум и максимум);
- разница между наибольшим и наименьшим значениями показателя (размах);
- сумма всех значений показателя, деленная на их количество (среднее арифметическое);
- выбросы (outliers) – значения показателей, резко отличающиеся основного числа значений.

1.3 Существующие инструментальные средства автоматизации при проведении административного мониторинга

Система обработки данных является частью АСУ и представляет собой комплекс инструментальных и организационных средств, предназначенный для обработки значений показателей объекта управления.

Информационная система с точки зрения законодательства РФ – это совокупность содержащейся в базах данных информации и обеспечивающих ее обработку информационных технологий и технических средств [33].

Информационные системы содержат модули получения, преобразования, хранения, аналитической обработки данных, формирования отчетов или генерации решений. Состав модулей определяется областью применения и назначением информационной системы.

В соответствии с характером обработки информации выделяют следующие типы информационных систем [34] [35]:

- информационные системы управления;
- системы аналитической обработки данных;
- системы интеллектуального анализа данных (data mining);
- инструменты конечного пользователя для выполнения запросов и построения отчетов (query and reporting tools).

1.3.1 Информационные системы управления

Данный вид систем ориентирован на тактический уровень управления, среднесрочное планирование и анализ. Для данного класса задач характерны регламентированность (периодическая повторяемость) формирования результатных документов и четко определенный алгоритм решения задач. Задачи решаются на основе накопленной базы оперативных данных [36].

1.3.2 Системы аналитической обработки данных и прочие средства многомерного анализа

OLAP (on-line analytical processing) – набор технологий для оперативной обработки информации, включающих динамическое построение отчётов в различных разрезах, анализ данных, мониторинг и прогнозирование ключевых показателей [37].

Основная идея OLAP заключается в построении многомерных таблиц, которые будут доступны для запросов пользователей. Согласно классификации OLAP-систем по способу хранения данных данные для многомерных таблиц могут храниться как в реляционных, так и многомерных базах данных. Поэтому применяются три способа хранения данных [38]: MOLAP (Multidimensional OLAP), ROLAP (Relational OLAP) и HOLAP (Hybrid OLAP).

Конкретный способ хранения удобен для конкретной задачи управления. Выбор способа управления определит требования к процедуре сбора данных, что означает изменение требований в зависимости от задачи управления.

1.3.3 Системы интеллектуального анализа данных (data mining)

Программные продукты, относящиеся к этой категории, обеспечивают получение качественно новой информации, не содержащейся в исходных данных [39].

1.3.4 Инструменты конечного пользователя для выполнения запросов и построения отчетов (query and reporting tools)

Системы данного класса предназначены для формирования запросов к информационным системам в пользовательских терминах, а также их исполнение, интеграцию данных из разных источников, просмотр данных с возможностями детализации и обобщения и построение полноценных отчетов, как экранных, так и печатных. Пользователь составляет запрос к источнику данных, используя заранее подготовленный программистом каталог терминов (семантический слой).

Инструменты конечного пользователя для выполнения запросов и построения отчетов поставляются двумя способами [40]:

- в составе OLAP-систем,
- в виде специализированных систем Query&Reporting.

Детальный анализ приведенных выше классов систем показывает то, что основной их функцией является аналитическая обработка данных и их обычно относят к классу информационно-аналитических систем. Для возможности их использования в процессах управления необходимо иметь подготовленный, определенным образом организованный массив данных.

1.4 Постановка задачи исследования

На основе анализа особенностей административного мониторинга средств и методов его организации можно заключить, что задача автоматизации процесса адаптации системы административного мониторинга является актуальной.

Детализируем основное направление и сформулируем основные способы проведения исследований.

Предлагается рассматривать систему мониторинга как составную часть автоматизированной системы управления производством производственного объединения, холдинга, корпорации. Следовательно, необходима формализация представления как объекта управления (в контексте задачи

организация административного мониторинга), так и процессов самой адаптивной системы административного мониторинга.

Требуется разработать формализованное представление задач получения, анализа и контроля первичных и сводных данных в терминах специальной модели хранения данных автоматизированной системы административного мониторинга, а также алгоритмы и информационное обеспечение средств адаптации автоматизированной системы административного мониторинга. Разрабатываемые модели и алгоритмы должны представлять собой инструментарий определения спецификаций и проектирования систем административного мониторинга, обеспечивающий адаптивность реализации процедур генерации отчетов, анализа и контроля данных, и, как следствие, снижение трудоемкости наиболее длительных по времени процессов адаптирующего и совершенствующего сопровождения.

1.5 Выводы по первой главе

1. Был проведен анализ особенностей административного мониторинга средств и методов его организации. По результатам анализа выявлено, что системам поддержки принятия решений и/или лицу принимающему решения, необходимо оперативно получать информацию из различных источников с последующей ее переработкой до необходимого уровня абстракции о состоянии объекта управления, в условиях изменяющихся требований к структуре и составу показателей. Вследствие чего должна обеспечиваться возможность оперативной адаптации автоматизированной системы административного мониторинга к этим требованиям.

2. Решение задачи оперативной адаптации автоматизированной системы административного мониторинга может быть достигнуто за счет формализованного представления задач получения, анализа и контроля первичных и сводных данных в терминах специальной модели хранения данных административного мониторинга.

3. Система мониторинга, в состав которой входят подсистемы генерации отчетов, анализа и контроля данных, соотнесена с классическим контуром управления в АСУ. Система мониторинга занимает промежуточное положение, получая данные об объекте управления непосредственно от самого объекта, выдает сводные данные, отличающиеся высоким уровнем абстракции, прошедшие контроль и анализ данных, затем передает информацию о состоянии объекта мониторинга системе, принимающей решения.

4. Введение адаптируемых программных средств системы административного мониторинга в АСУП приводит к появлению контура управления самой системой мониторинга. Данный контур обеспечивает адаптацию процедур получения сводных данных, анализа и контроля данных в соответствии с требованиями задачи административного мониторинга.

5. Вследствие уникальности модели хранения данных на физическом уровне, реализованной в адаптивной системе административного мониторинга, ни один из существующих генераторов отчетов и ни одна из технологий генерации не могут быть напрямую перенесены в рассматриваемую систему. Однако, базовые принципы на которых основывается генерация отчета в системах различных классов могут быть (с определенной модификацией) использованы при построении подсистемы генерации отчетов адаптивной системы административного мониторинга.

6. Для генерации простых отчетов должны быть сформированы модельные представления шаблона отчета, структуры отчета и правил его формирования. На основе данных модельных представлений возможна автоматическая генерация отчетной формы.

2 Формализация задач получения, анализа и контроля первичных и сводных данных при проведении административного мониторинга

2.1 Модель процесса переработки данных

В ходе анализа задач получения, анализа и контроля первичных и сводных данных была разработана модель процесса переработки данных [5]. Ее можно представить помощью графа (рисунок 2.1).

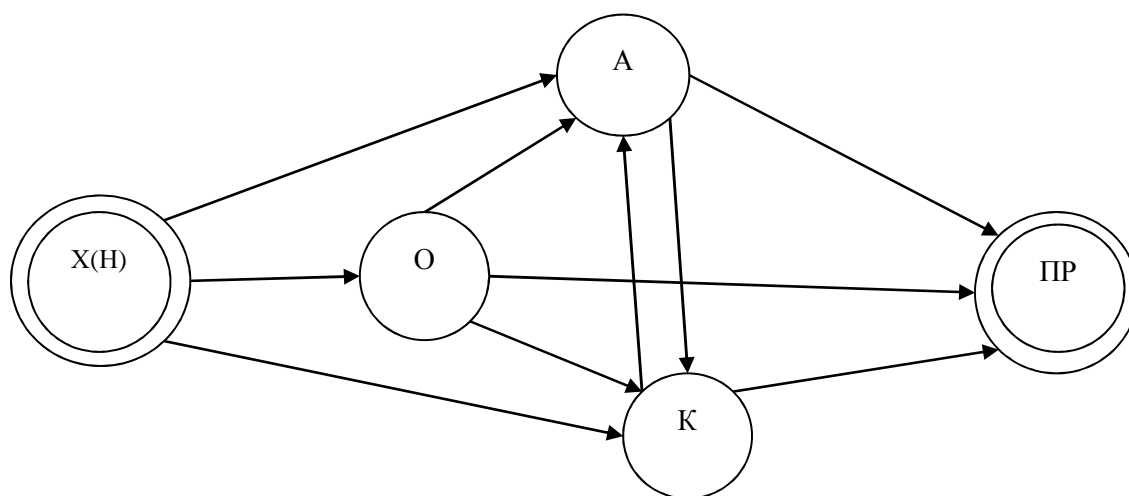


Рисунок 2.1 – Модель процесса переработки данных

Источком является процесс накопления (X(H)) данных объекта управления, стоком является процесс принятия решения (PP). Дуга (X(H),O) показывает множество экземпляров объектов учета и их показателей, которые процесс генерации отчетов (O) получает от процесса накопления. Процесс анализа данных (A), в зависимости от вида задачи анализа данных, получает от процесса накопления (дуга (X(H),A)) совокупность экземпляров объектов учета одного типа с выбранными показателями, принадлежащими данному типу, или экземпляр объекта учета со значениями периодического показателя, принадлежащего данному экземпляру объекта учета. Процесс контроля данных (K) получает от процесса накопления (дуга (X(H),K)) множество экземпляров объектов учета одного типа с показателем,

принадлежащим данному типу. От процесса генерации отчетов процессы принятия решений (ПР), анализа и контроля данных (дуги (О,ПР), (О,А) и (О,К)) получают совокупность отношений между собой множеств экземпляров ОУ с итоговыми показателями (полученные путем агрегирования), характеризующими состояние объекта управления. От процесса анализа данных процессы контроля данных и принятия решений получают (дуги (А,К) и (А,ПР)), в зависимости от задачи анализа данных, множество экземпляров объектов учета, разбитых по разным группам, или экземпляр объекта учета со значениями периодического показателя для будущих периодов учета. От процесса контроля данных процессы анализа данных и принятия решения получают (дуги (К,А) и (К,ПР)) множество экземпляров объектов учета разбитых на разные группы. Основываясь на формализме представления объекта мониторинга в терминах модели хранения данных, приведенной выше, можно вывести следующее формализованное представление информационных потоков:

Дуга (X(H),O): V_o, P ,

где V_o – множество экземпляров объектов учета, P – множество показателей.

Дуга (X(H),A): $V_{ot}, X_{pt1}, o_1, p_1, S_n$,

где V_{ot} – множество объектов учета типа t , X_{pt1} – подмножество набора показателей ассоциированных с типом t , $o_1 \in V_o$ – экземпляр объекта учета, имеющий периодический показатель p со множество значений S_n за n периодов учета.

Дуга (X(H),K): V_{ot}, p_1 ,

где V_{ot} – множество объектов учета типа t , имеющих показатель p_1 .

Дуги (О,ПР), (О,А), (О,К): $V_o, P' = f(P)$,

где P' – множество итоговых показателей, полученных в результате применения функции f ко множеству показателей P . Функция f может являться функцией агрегирования, среднего арифметического, сложения по

модулю, выбора минимального, выбора максимального в зависимости от задачи мониторинга.

Дуги (А,К), (А,ПР) – $V_{oz}, o_1, p_1, S_{n+k}$,

где V_{oz} – множество z групп (классов, кластеров, подмножеств) экземпляров объектов учета, $o_1 \in V_o$ – экземпляр объекта учета, имеющий периодический показатель p со множество значений S_{n+k} за $n+k$ периодов учета.

Дуги (К,А), (К,ПР) – V_{ox} ,

где V_{ox} – множество x групп (подмножеств) экземпляров ОУ.

2.2 Методическое обеспечение автоматизации процессов получения, анализа и контроля первичных и сводных данных

2.2.1 Методика настройки и получения сводных данных административного мониторинга

На основе сформулированных в первом разделе принципов построения подсистемы генерации отчетов может быть сформулирована методика, обеспечивающая автоматизированную реализацию процессов настройки и генерации выходных отчетных форм в системе административного мониторинга.

Графическое представление методики приведено на рисунке 2.2.

После запуска подсистемы генерации выходных отчетных форм возможны три сценария использования подсистемы:

1 Формирование выходной отчетной формы на основе нового (создаваемого) шаблона отчета.

2 Формирование выходной отчетной формы на основе созданного ранее шаблона отчета.

3 Формирование выходной отчетной формы на основе одного из типовых шаблонов, предустановленных в подсистеме.



Рисунок 2.2 – Методика настройки и генерации выходных отчетных форм

В случае реализации первого сценария происходит создание нового стиля визуализации, то есть определение модели структуры и правил формирования выходной отчетной формы. Затем создается шаблон отчета, определяющий его информационное содержание. После сохранения полученного шаблона и стиля визуализации возможна непосредственная генерация отчета, осуществляемая при запуске визуализатора на основе созданных шаблона и стиля визуализации.

Второй сценарий отличается от первого лишь тем, что первым шагом является выбор созданного ранее шаблона и, соответственно, связанного с ним стиля визуализации.

В случае выбора пользователем третьего сценария осуществляется выбор одного из предустановленных в системе шаблонов, в котором уже определена абстрактная информационная структура отчета и определен стиль визуализации. Данный шаблон интерпретируется с целью определения множеств исходных данных, которые требуется конкретизировать (типы объектов учета, конкретные объекты учета, показатели). На основе полученного множества осуществляется автоматическая генерация [41] специализированного диалогового интерфейса ввода данных, на котором пользователь имеет возможность конкретизировать множество исходных данных для построения данного типового отчета. Процедура же генерации (визуализации) типовой отчетной формы не отличается от процедуры генерации на основе созданных шаблонов отчетов.

2.2.2 Методика построения процессов автоматизации анализа и контроля данных

На основании модели процесса переработки данных была разработана методика построения процессов автоматизации анализа и контроля данных в системе административного мониторинга (рисунок 2.3). Она включает три основных этапа.

1 Задание правил и критериев анализа и контроля данных. В рамках данного этапа происходит инициализация определяемых пользователем переменных, являющихся аргументами функций, полученных при моделировании анализа и контроля. Также определяются общие настройки задачи анализа и контроля, которые в совокупности с аргументами функции формируют общую модель задачи анализа и контроля.

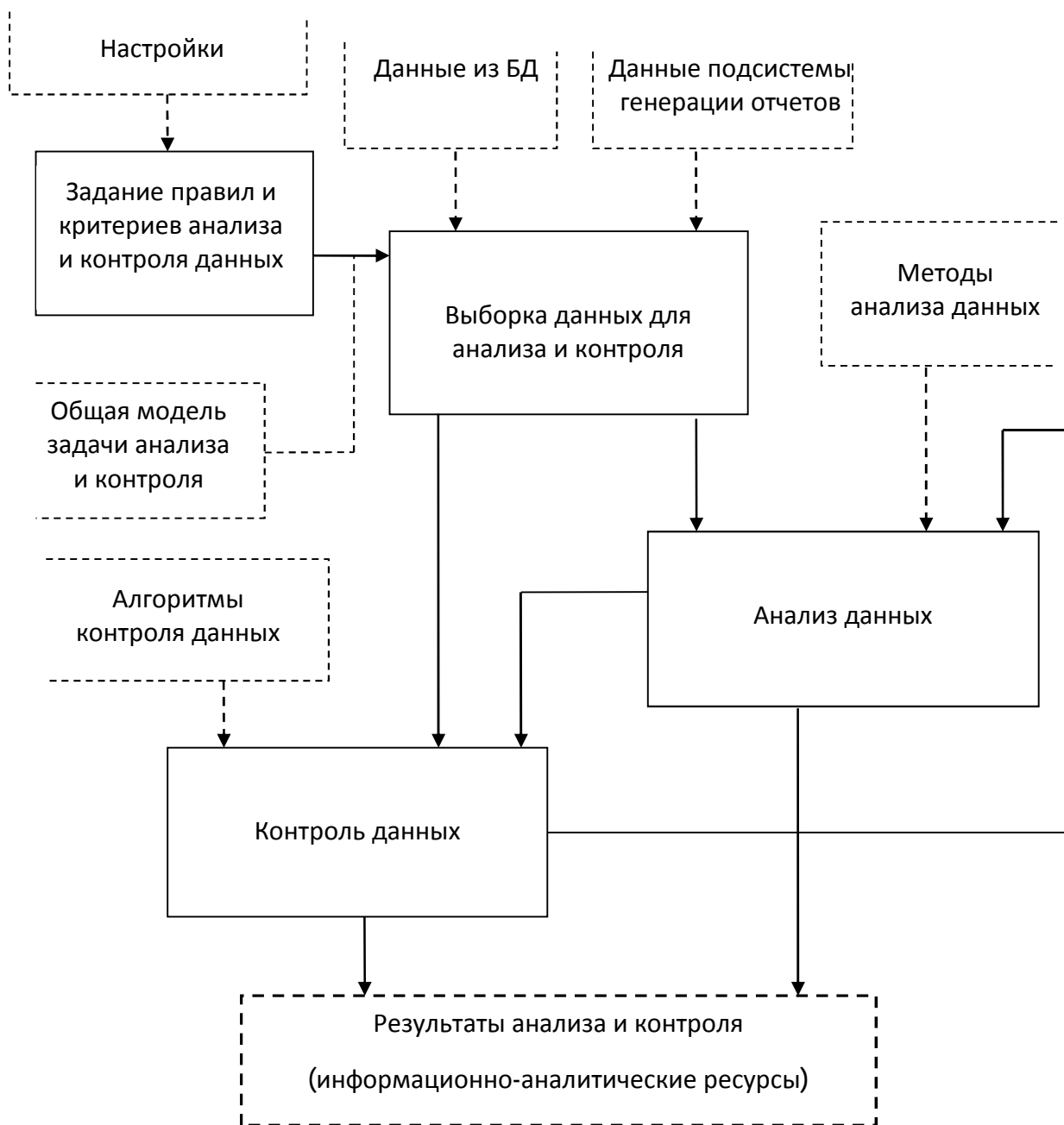


Рисунок 2.3 – Методика построения процессов автоматизации анализа и контроля данных

2 *Выборка данных для анализа и контроля.* На данном этапе на основе общей модели задачи анализа и контроля формируется соответствующий конкретной задаче исходный набор данных. Выборка данных может осуществляться двумя способами в зависимости от определенного в модели задачи источника данных:

– непосредственная выборка из базы данных (задача анализа и/или контроля решается на первичных показателях);

– получение данных из подсистемы генерации отчетов (задача анализа и/или контроля решается на сводных данных, прошедших первичную обработку).

3 Анализ и контроль данных. Данный этап может быть разбит на два подэтапа, реализуемых последовательно или по отдельности.

3.1 Анализ данных. Анализ данных в любом случае осуществляется на основании данных, полученных по результатам выполнения второго этапа. Его результаты в подавляющем большинстве случаев могут рассматриваться как самостоятельные информационно-аналитические ресурсы, пригодные для принятия управленческих решений. Также возможно использование этих результатов как объекта контроля, тем самым повышая уровень абстракции ИАР. Инструментарием реализации данного подэтапа являются методы анализа данных, выбираемые пользователем в зависимости от специфики исходных данных методы анализа данных, которые должны быть реализованы в виде библиотеки процедур в составе системы административного мониторинга.

3.2 Контроль данных. Исходными данными для контроля данных могут являться результаты этапа 2 или подэтапа 3.1. Результаты контроля данных могут использоваться как информационно-аналитические ресурсы, пригодные для принятия управленческих решений. Также возможно использование этих результатов как объекта анализа, тем самым повышая уровень абстракции ИАР. Инструментарием при реализации данного подэтапа являются реализованные в программных модулях системы административного мониторинга алгоритмы контроля данных.

2.2.3 Общая модель анализа и контроля данных

Для задачи анализа и/или контроля данных должны определяться ограничения на исходное множество данных, источник данных и настройки

запуска процедуры и отображения результатов и др. Поэтому общая модель задачи анализа и/или контроля данных должна содержать (определять):

- название задачи (символьный, имеющий значение для пользователя, возможно, отражающий цель применения, идентификатор задачи анализа и/или контроля);
- номер (идентификатор) правила анализа или критерия контроля данных;
- режим запуска (по расписанию или вручную);
- запрос выборки данных (отражает способ получения исходных данных для анализа и/или контроля: выборка из базы данных, данные отчетов, данные анализа данных);
- период выборки данных (используется по умолчанию в случае запуска по расписанию; при запуске вручную может быть изменен);
- связь с выходным интерфейсом (идентификатор диалогового интерфейса вывода данных, при обращении к которому режиме осуществляется решение задачи (при ручном режиме) и/или отображение результатов).

В итоге общую модель анализа и контроля данных можно представить следующим кортежем (обозначения атрибутов приведены в порядке приведенного выше перечисления):

Model=<name, number, mode, request, period, interface>.

2.3 Моделирование структуры, правил формирования и содержимого выходных отчетных форм на основе шаблонов

Представление шаблона отчета – дерево, каждый уровень которого определяет количество потомков до корневого узла. Узлами данного дерева являются объекты учета, которые могут быть представлены только названием, а могут нести с собой массив (или массивы, для периодических показателей) данных со значениями показателей [42].

Возможность привязывать к узлам структуры отчета массивы данных со значениями или нет позволяет генерировать как отчеты, требующие представления непосредственно мониторинговых данных, так и отчетов, требующих представления только связей между объектами учета.

Аналитическую обработку данных возможно производить как «горизонтально», так и «вертикально». Горизонтальная обработка подразумевает группировку и применение агрегирующих и других аналитических функций к каждому показателю для всех объектов учета одного поколения. Возможна также подгруппировка в пределах одного уровня – по типам.

Вертикальная обработка значений числовых показателей подразумевает применение агрегирующих и других аналитических функций для каждого показателя к сгруппированным объектам учета, принадлежащим одному родителю. Промежуточные аналитические данные фиксируются на каждом уровне структуры, а затем группируются по всем уровням. Вертикальная обработка данных возможна, только если выбраны одинаковые показатели на нескольких уровнях. В противном случае интерфейс с указанием правил вертикальной аналитики не будет доступен.

2.3.1 Язык описания шаблонов выходных отчетных форм

Для получения сводных данных в виде отчетов, был разработан язык описания шаблона выходных отчетных форм, который определяет модель структуры и содержания выходных отчетных форм.

Для определения языка описания шаблона воспользуемся формальной системой описания искусственных языков, в которой одни синтаксические категории последовательно определяются через другие. Применим форму Бэкуса-Наура [43], поскольку она позволяет описывать контекстно-свободные грамматики. Как в любой контекстно-свободной грамматике существуют как терминальные, так и нетерминальные символы. Теги, начинающиеся со слов «Показать» являются терминальными элементами,

которые при формировании конечной формы отчета поступают на вторую стадию генерацию отчетов – работу визуализатора с полученной структурой. Остальные элементы – нетерминальные, которые преобразуются дальше с использованием правил продукции [42].

Формальное описание шаблона отчета следующее.

<уровень> ::= “{” <тип уровня>+ “}”.

<тип уровня> ::= <Показать имя типа> “:”

<данные типа>.

<данные уровня> ::= “{” <объекты учета>,

<показатели типа>,

<горизонтальный анализ>,

<вертикальный анализ>,

<потомки> “}”.

<показатели типа> ::= “properties: [

” (<Показать имя показателя> “,”)* < Показать имя показателя > “]”.

<горизонтальный анализ> ::= “horizont_analis:

{” <пустой список> |

((<функция на показателе> “,”)*

< функция на показателе >) “}”.

<пустой список> ::= “empty”.

<функция на показателе> ::=

<Показать имя показателя> “:” <функция>.

<функция> ::= “SUM” |

“MAX” |

“MIN” |

“AVG” |

“XOR”.

<вертикальный анализ > ::= “vert_analis:

” <пустой список> |

((<функция на показателе с правилом «подъема»> “,”)* < функция на показателе с правилом «подъема»>) “}”.

<функция на показателе с правилом «подъема»> ::=

<Показать имя показателя> “:”

<функция с правилом «подъема»>.

<функция с правилом «подъема»> ::=

<Показать имя функции>&<правило>.

<правило> ::=

<Показать правило, учитывающее значение 1 раз> | <Показать правило, учитывающее все значения> | <Показать правило, пропорционально учитывающее значения>.

<потомки> ::= <уровень>.

Необходимо отметить, что набор значений тегов «функция» и «правило» при необходимости введения новых функций или правил обработки могут расширяться. Их реализация должна осуществляться в виде отдельных программных элементов, пригодных для включения новых программных функций.

2.3.2 Пример описания отчета с использование предложенного языка

Покажем применение приведенного выше языка на примере построения отчета о количестве методических указаний (МУ) разработанных в университете (в целом и по подразделениям) сотрудниками входящих в него подразделений («Факультет 1», «Факультет 2», Центр). Также в отчете выводятся данные по подразделениям (сумма отработанных сотрудниками часов, максимальная зарплата) и детальные показатели по сотрудникам.

report1 =

{ objects: [«Университет1»]

properties: «»,

vert_analis:

```
{«МУ»: SUM  
},
```

childs:

```
{ objects: [ «Факультет 1»,  
             «Факультет 2»,  
             «Центр1»  
           ],
```

properties: «»,

vert_analis:

```
{«МУ»: SUM},
```

childs:

```
{ objects: [ «Сотрудник 1»,  
             «Сотрудник 2»,  
             «Сотрудник 3»  
           ],
```

properties: [«должность»,

«тел»,

«КОЛ-ВО ЧАСОВ»,

«МУ»,

«зарплата»

],

vert_analis:{empty},

horiz_analis:

```
{«КОЛ-ВО ЧАСОВ»: SUM,
```

```

        «МУ»: SUM,
        «зарплата»:MAX
    },
    childs:{empty}
},
},
}.

```

2.3.3 Модель структуры отчета и правил его формирования

Как было сказано выше шаблон отчета задает его информационную структуру, то есть определяет множество данных, отображаемых на выходной отчетной форме. Шаблон отчета в том числе является абстрактной моделью древовидной структуры данных, которая будет использоваться в процессе выборки данных из базы данных системы административного мониторинга. Получаемая таким образом информация (в случае сложной структуры объекта мониторинга) может быть выведена на выходную форму и визуальное структурированное и группирована различными способами. Параметры структуризации и группировки информации, ее отображения определяются моделью структуры отчета и правил его формирования. Поэтому, учитывая тот факт, что данная модель не определяет непосредственно данные, а регламентирует процесс отображения данных, представим ее в виде стиля визуализации отчета.

Стиль визуализации отчета можно укрупненно разделить на три основных компонента:

- стиль представления данных экземпляров объектов учета;
- стиль консолидации представлений (агрегированного представления) экземпляров объектов учета (с данными или без них);

– стиль представления результатов аналитических функций (функций обработки данных).

Стиль визуализации определяет выбор правил преобразования структуры шаблона. В реализации данное преобразование представляет собой получение текстовой разметки из структурированного текста (структуры шаблона) с использованием атрибутов стиля согласно алгоритму, реализующему правила преобразования. Упрощенная схема процесса генерации отчета приведена на рисунке 2.4.



Рисунок 2.4 – Упрощенная схема процедуры генерации выходной отчетной формы

Результат выбора стиля в каждом компоненте является значением одного параметра. Значения трех указанных параметров и будут представлять собой «единицу» хранения стиля визуализации. Создание стиля визуализации отчета осуществляется за счет последовательного определения всех трех параметров. Обозначим указанные атрибуты для краткости и возможности обращаться к ним в формальной записи, а впоследствии – в алгоритме:

- `dataStyle` – стиль представления данных экземпляров объектов учета;
- `consolidationStyle` – стиль консолидации представлений экземпляров объектов учета;
- `analysisStyle` – стиль представления результатов аналитических функций.

Рассмотрим выделенные элементы стиля визуализации отчета подробнее.

2.3.3.1 Стиль представления данных экземпляров объектов учета

Настройка представления начинается с определения стиля визуализации узла структуры шаблона, то есть экземпляра объекта учета с возможным ассоциированным массивом данных административного мониторинга. В случае если показ данных не предусмотрен структурой шаблона, то этот шаг будет пропущен. Узел структуры шаблона будет представлен одним названием экземпляра объекта учета.

Если показ данных мониторинга у узла предусмотрен, то возможны следующие представления:

- графическое;
- табличное.

Графическое представление экземпляра объекта учета предусматривает построение графика, либо гистограммы. Под графиком будем подразумевать тип диаграмм, на которых полученные данные изображаются в виде точек, соединенных или не соединенных соединительными линиями. Для построения графиков используют прямоугольную систему координат. На графиках возможно показать связь только двух источников. Как правило, это либо график значений периодических показателей для указанных по оси абсцисс периодов учета, либо это обобщенный график для определенного показателя, показывающий его значения (например, среднее, минимальное, максимальное) для объектов учета одного типа.

Под гистограммами будем подразумевать столбчатые или линейные диаграммы, изображающие данные в виде вертикальных или горизонтальных прямоугольников. В гистограммах возможно показать связь как двух, так и трех источников. Гистограмма, строящаяся для экземпляра объекта учета источником данных оси абсцисс может иметь периоды учета определенного

показателя (см. пример на рисунке 2.5), либо сгруппированные показатели для нескольких объектов учета (см. пример на рисунке 2.6). В последнем случае речь идет о сгруппированных столбчатых диаграммах. По оси ординат располагаются значения указанных показателей (показателя).

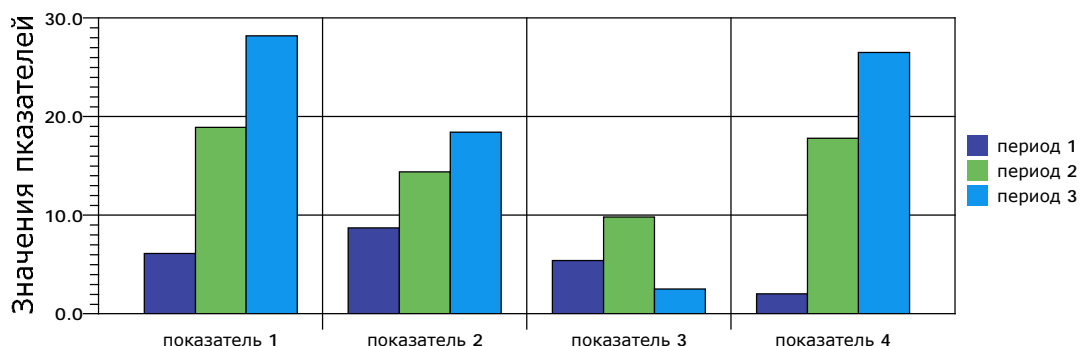


Рисунок 2.5 – Пример гистограммы изменения значений показателей для заданного объекта учета

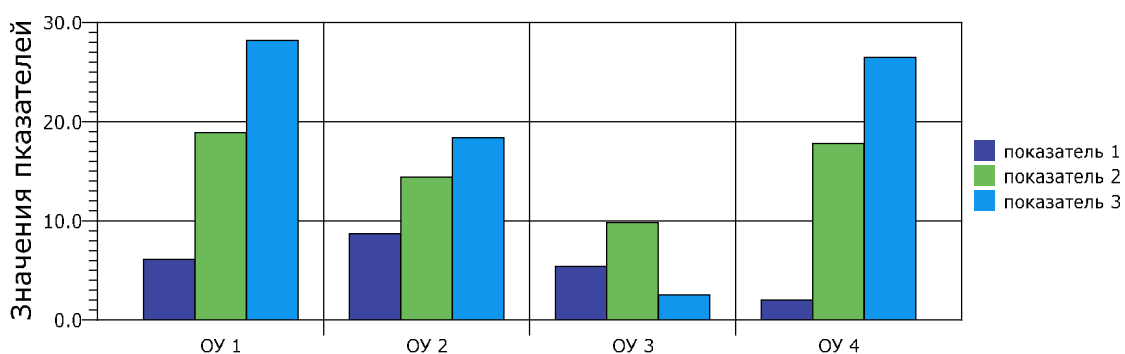


Рисунок 2.6 – Пример гистограммы представления значений показателей для нескольких объектов учета

В первом случае атрибут `dataStyle` примет следующий вид:

```
dataStyle : {
    type : "graph",
    x : "durations",
    xGroup : "properties"
}
```

Во втором случае:

```
dataStyle : {
```

```
    type : "graph",
    x : "properties",
    xGroup : "objects"
}
```

Также необходимо отметить, что возможны и другие комбинации источников данных.

Табличное представление данных экземпляра объекта учета предполагает построение:

- таблицы со значениями показателей в ячейках и названиями показателей в заголовках (для общих показателей);
- таблиц со значениями показателей за определенный период в ячейках и периодами в заголовках (для периодических показателей).

В связи с этим необходимо ввести отдельное поле в атрибуте `dataStyle`, определяющее является ли рассматриваемая группа показателей общими или периодическими – `propertyKind`.

Кроме того табличное представление данных экземпляра ОУ можно транспонировать, что определит расположение заголовков. Стиль отображения заголовка определяется полем «`direction`» в атрибуте и может принимать значения: горизонтальный или вертикальный, что показывает расположение набора показателей.

В случае представления данных по периодическим показателям возможно производить визуальное совмещение временных шкал с указанием периодов действия значения показателя для нескольких показателей. Альтернативным решением является отображение наборов значений периодических показателей в отдельных таблицах, без единой временной шкалы. Данное свойство определяется в поле `synchronTimeLine`.

Таким образом, атрибут `dataStyle` примет вид.

```
dataStyle : {
    type : "table",
    propertyKind : common | periodical,
```

```
direction : "horizont" | "vertical",  
[ synchronTimeLine : yes | no ]  
}
```

2.3.3.2 Стиль консолидации представлений экземпляров объектов учета

Стиль консолидации представлений экземпляров объектов учета определяет способ объединения представленных данных экземпляра учета. Если показ данных не предусмотрен структурой шаблона, то на данном шаге производится настройка стиля визуализации иерархии экземпляров объектов учета.

Выбор правил визуализации иерархии экземпляров объектов учета. Иерархия экземпляров объектов учета представима несколькими способами. По результатам предварительного анализа принято решение использовать два наиболее наглядных: матрица смежности и список смежности (часто представляемый в виде дерева).

Матрица смежности представляет собой таблицу, где столбцы соответствуют объектам учета одного уровня, а строки соответствуют объектам учета соседнего уровня. В каждой ячейке этой матрицы записывается маркер, определяющей наличие родительской связи от объекта учета в строке к объекту учета в столбце (либо наоборот).

Матрица смежности удобна для представления связей двух соседних поколений, но не для представления большего числа уровней. Например, матрицей смежности удобно пользоваться для визуализации отчета о распределении сотрудников по отделам.

Атрибут, отвечающий за правила объединения представлений данных экземпляров объектов учета, примет следующий вид:

```
conslidationStyle: {  
    type : "hierarch",  
    direction : "horizont" | "vertical"
```

```
}
```

Поле «direction» определяет ось, по которой будут перечислены объекты учета родительского уровня. Если в нем указано «horizont», то родительские экземпляры объектов учета будут давать названия столбцов.

В случае выбора древовидного представления, conslidationStyle будет иметь вид:

```
consolidationStyle: {  
    type : "tree"  
}
```

Выбор правил визуализации объединения представлений данных экземпляров объектов учета.

Если структурой шаблона показ данных предусмотрен, то на шаге определения стиля визуализации объединения этих данных можно выбрать варианты:

- показать объединенные данные в единой таблице;
- показать объединенные данные в виде дерева таблиц;
- показать семейство диаграмм (если данные мониторинга представляются графическим способом).

Объединенная таблица представляет собой таблицу, содержащую значения показателей всех объектов учета одного типа одного родителя (если в аналитических функциях планируется обработка данных с агрегацией по типам) или всех объектов учета одного родителя, без разбивки на отдельные таблицы по типам. Заголовок с названиями показателей в данном случае один.

Если таблицы строятся для последнего рассматриваемого уровня, то визуализация очевидна, поскольку таблица не прерывается детализацией некоторых объектов на следующем уровне. Если же таблицы строятся для любого уровня, то вложенные уровни должны представлять собой особые строки в таблице с возможностью визуального сокрытия.

При этом отдельно объединяются значения периодических показателей, отдельно – общих.

Объединение данных в виде дерева может быть представлено как семейство таблиц, привязанных к названию экземпляра объектов учета с изображенными графически связями с родительским экземпляром объекта учета.

Объединение диаграмм значений показателей в семейство выглядит аналогично – с изображенными связями с родительским экземпляром объекта учета и возможностью визуального сокрытия диаграмм уровня.

Опишем атрибут `consolidationStyle` для подобного случая:

```
consolidationStyle: {  
    type : "unionTable" | "treeOfTables" |  
        "treeOfGraph",  
    unionFactor : "type&parent" | "parent"  
}
```

2.3.3.3 Стиль представления результатов аналитических функций

Аналитические функции обрабатывают информацию двух видов: количество объектов учета, состоящих в тех или иных связях с остальными объектами учета и значения количественных показателей объектов учета (как общих, так и периодических).

Агрегирующая обработка данных может иметь источником:

- набор значений показателей объектов учета, принадлежащих одному родителю и одного типа;
- набор значений показателей объектов учета одного родителя;
- набор значений показателей объектов учета всего уровня.

Эти наборы данных можно рассматривать как вложенные, в том смысле, что обрабатывать можно уже агрегированные данные. Так агрегированные значения показателей объектов учета, принадлежащих одному родителю, можно обрабатывать повторно для получения

обобщенных данных об объектах всего уровня. Вложенные источники агрегации можно пропускать, о чем будет свидетельствовать атрибут «aggregation».

Общий вид атрибута analysisStyle следующий:

```
analysisStyle : {  
    info : "count" | "values",  
    aggregation : [ ["type&parent",] "parent"],  
    "level".  
}
```

2.4 Моделирование критериев контроля и правил анализа данных

2.4.1 Формализация представления задач анализа и контроля данных в системе административного мониторинга

Среди задач анализа данных выделим, как наиболее часто востребованные в практической деятельности, следующие задачи [32]:

- классификация;
- кластеризация;
- прогнозирование.

2.4.1.1 Формализация задачи классификации

Процесс классификации заключается в разбиении множества объектов учета на классы по определенному критерию. В терминах модели представления объекта мониторинга можно сказать, что в задаче классификации требуется определить значение зависимого показателя объекта учета на основании значений других показателей, характеризующих данный объект. В качестве анализируемых показателей могут выступать любые наблюдаемые, представленные в базе данных системы мониторинга

первичные показатели. В качестве зависимого показателя будет выступать свойство объекта быть отнесенным к одному из определенных аналитиком классов.

Формально постановка задачи выглядит следующим образом [32]. Есть множество объектов учета V_o . Каждый o_j объект из множества V_o характеризуется набором показателей X_{pj} .

$$X_{pj} = \{p_1, p_2, \dots, p_h, \dots, p_m, y\},$$

где p_h – независимые показатели, значения которых известны и на основании которых определяется зависимый показатель y . Множество значений S , которые может принимать показатель y , заранее определено и конечно.

$$S = \{s_1, s_2, \dots, s_r, \dots, s_k\},$$

где k – количество классов, на которые разбиваются исследуемые объекты учета.

В задаче классификации обнаруженная функциональная зависимость может быть представлена с помощью математической функции. В этом случае исследуемые объекты учета будут рассматриваться как точки в $(m+1)$ -мерном пространстве. Тогда набор показателей X_{pj} объекта учета o_j представляются как координаты, а функция имеет следующий вид:

$$y_j = \omega_0 + \omega_1 p_1 + \omega_2 p_2 + \dots + \omega_m p_m,$$

где $\omega_0, \omega_1, \dots, \omega_m$ – веса независимых показателей, в поиске которых и состоит задача нахождения классификационной функции. При этом все категориальные и логические показатели должны быть преобразованы к числовым.

2.4.1.2 Формализация задачи кластеризации

Задача кластеризации состоит в разделении исследуемых объектов учета на группы «похожих» объектов, называемых кластерами. В отличие от задачи классификации в задаче кластеризации отнесение каждого из объектов учета системы административного мониторинга осуществляется к

одному (или нескольким) из заранее неопределенных классов. Разбиение объектов учета по кластерам осуществляется при одновременном их формировании. Определение кластеров и разбиение по ним объектов учета выражается в итоговой модели данных, которая является решением задачи кластеризации.

Формально задача кластеризации описывается следующим образом [32]. Есть множество исследуемых объектов учета V_o . Требуется построить множество кластеров C и отображение F множества V_o на множество C . Отображение F задает модель данных, являющуюся решением задачи:

$$F: V_o \rightarrow C,$$

$$C = \{c_1, c_2, \dots, c_k, \dots, c_g\},$$

где c_k – кластер, содержащий «похожие» друг на друга объекты из множества V_o :

$$c_k = \{o_j, o_i \mid o_j \in V_o \& o_i \in V_o \& d(o_j, o_i) < \sigma\},$$

где σ – величина, определяющая меру близости для включения объектов в один кластер; $d(o_j, o_i)$ – мера близости между объектами, называемая расстоянием.

Значение $d(o_j, o_i)$ называется расстоянием между объектами учета o_j и o_i , если выполняются следующие условия:

- 1 $d(o_j, o_i) > 0$, для всех o_j и o_i .
- 2 $d(o_j, o_i) = 0$, тогда и только тогда, когда $o_j = o_i$.
- 3 $d(o_j, o_i) = d(o_i, o_j)$.
- 4 $d(o_j, o_i) \leq d(o_j, o_r) + d(o_r, o_i)$.

Если расстояние $d(o_j, o_i)$ меньше некоторого значения σ , то говорят, что элементы близки и помещаются в один кластер. В противном случае говорят, что элементы отличны друг от друга и их помещают в разные кластеры.

2.4.1.3 Формализация задачи прогнозирования

Прогнозирование в системе административного мониторинга направлено на определение тенденций динамики конкретного объекта учета

(как элементарного, так и составного) на основе ретроспективных данных, т.е. анализа его состояния в прошлом и настоящем. Таким образом, решение задачи прогнозирования требует некоторой обучающей выборки данных. Задачу прогнозирования можно сформулировать как установление функциональной зависимости между зависимыми и независимыми показателями объектов учета. Различие задач классификации и прогнозирования состоит в том, что в первой задаче предсказывается класс зависимого показателя объекта учета, а во второй – числовые значения зависимого показателя объекта учета, пропущенные или неизвестные (относящиеся к будущему).

Основой для прогнозирования в случае мониторинга (ввиду его проведения на периодической основе) служит историческая информация, хранящаяся в базе данных в виде временных рядов. Временной ряд – последовательность наблюдаемых значений какого-либо периодического показателя исследуемого объекта учета, упорядоченных в неслучайные моменты времени. Обозначим генеральную совокупность, которой принадлежат все значения показателей через S . Таким образом, S представляет собой поток значений, поступающих в систему мониторинга извне. После контроля и первичной обработки значение подается на вход подсистемы хранения. Преобразованный поток значений обозначим через V_s . Периодические показатели, то есть показателей с переменными значениями в зависимости от того, значение какого периода введено, формально выглядят в модели как наборы значений. Данный набор представляет собой множество упорядоченных пар: период – значение за период.

Таким образом,

$$V_s = \{d, s_i \mid d \in D \ \& \ s_i \in S\},$$

где d – идентификатор периода учета значений показателя s .

Значения периодических показателей обычно фиксируются через равные промежутки времени и могут быть представлены в виде последовательности $\{s_1, s_2, \dots, s_i, \dots, s_n\}$, где n – количество периодов учета.

Задачу построения прогноза по временному ряду можно сформулировать следующим образом [32]. Имеются значения периодического p_h показателя объекта учета o_j за n периодов:

$$p_h = \{s_1, s_2, \dots, s_i, \dots, s_n\},$$

На их основании требуется определить значение s_{n+k} , при $k > 0$.

2.4.1.4 Формализация задач контроля данных

Задача контроля данных состоит в сравнении значений показателей, характеризующих состояния объектов учета со значениями определенных параметров и их оценку на соответствие этим параметрам. Такими параметрами являются [32]:

- наименьшее и наибольшее значения показателя (минимум и максимум);
- разница между наибольшим и наименьшим значениями показателя (размах);
- сумма всех значений показателя, деленная на их количество (среднее арифметическое);
- выбросы (outliers) – значения показателей, резко отличающиеся основного числа значений.

Задачу контроля наименьшего и наибольшего значения показателя можно сформулировать следующим образом следующим образом: имеется множество всех показателей P свойственных объектам учета определенного типа и множество S_i , содержащее значения конкретного показателя p_i :

$$S_i = \{s_{i1}, s_{i2}, \dots, s_{ij}, \dots, s_{ik}\}$$

Необходимо сформировать множество P_m .

$$P_m = \{m_1, m_2, \dots, m_i, \dots, m_n\},$$

где m_i – показатель, прошедший контроль наименьшего и наибольшего значения:

$$m_i = \{p_i \mid p_i \in P \ \& \ s_{ij} \in S_i \ \& \ s_{ij} < \max_i \ \& \ s_{ij} > \min_i\},$$

где \max_i и \min_i – задаваемые в процессе подготовки задачи контроля минимальное и максимальное допустимые значения показателя p_i .

Сформулируем задачу контроля разницы между наибольшим и наименьшим значением показателя. На основе исходных данных предыдущей задачи необходимо сформировать множество P_r :

$$P_r = \{ r_1, r_2, \dots, r_i, \dots, r_n \},$$

где r_i – показатель, прошедший контроль разницы между максимальным и минимальным значением:

$$r_i = \{ p_i \mid p_i \in P \ \& \ s_{i\max} \in S_i \ \& \ s_{i\min} \in S_i \ \& \ |s_{i\max} - s_{i\min}| < \text{int}_i \},$$

где int_i – задаваемая в ходе настройки допустимая разница между наибольшим и наименьшим значениями p_i показателя;

$s_{i\max}$ и $s_{i\min}$ – наибольшие и наименьшие значения показателя p_i .

Задача контроля среднего арифметического формулируется следующим образом: необходимо сформировать множество P_a :

$$P_a = \{ a_1, a_2, \dots, a_i, \dots, a_n \},$$

где a_i – показатель, прошедший контроль среднего арифметического:

$$a_i = \{ p_i \mid p_i \in P \ \& \ \text{avr}(k, S_i) < a_{i\max} \ \& \ \text{avr}(k, S_i) > a_{i\min} \},$$

где k – количество значений показателя p_i ;

$\text{avr}(k, S_i) = \frac{1}{k} \sum_{j=1}^k s_{ij}$ – среднее арифметическое значений p_i показателя;

$a_{i\max}$ и $a_{i\min}$ – задающиеся минимальное и максимальное допустимые значения среднего арифметического значений p_i показателя.

Сформулируем задачу контроля выбросов – значений показателей, резко отличающихся от основного числа значений. Необходимо сформировать множество P_g :

$$P_g = \{ g_1, g_2, \dots, g_i, \dots, g_n \},$$

где g_i – показатель прошедший контроль выбросов:

$$g_i = \{ p_i \mid p_i \in P \ \& \ s_{i, h+1} \in S_i \ \& \ h = \overline{1, k-1} \ \& \ |\text{avr}(h, S_i) - s_{i, h+1}| < \text{out} \},$$

где k – количество значений показателя p_i ;

$$\text{avr}(h, S_i) = \frac{1}{h} \sum_{j=1}^h s_{ij} - \text{среднее арифметическое значений показателя } p_i;$$

out – задаваемая допустимая разница между средним арифметическим значений показателя p_i и рассматриваемым значением $s_{i, h+1}$.

2.4.2 Модели правил анализа данных

На основе приведенных в предыдущем разделе формальных постановок задач анализа и контроля необходимо построить формализованные модели правил анализа и критериев контроля данных [44]. Данное моделирование преследует две цели. Во-первых, необходимо в терминах модели представления объекта административного мониторинга формализовать рассмотренные задачи анализа и контроля. То есть определить информационную модель задач, которая будет использоваться в качестве основы при проектировании как внутренних структур данных подсистемы анализа и контроля, так и структуры информационных потоков, связывающих подсистему хранения с подсистемой анализа и контроля. Во-вторых, будет получено абстрактное представление процедур преобразования данных, осуществляемых в ходе решения задач анализа и контроля. То есть будут получены спецификации процедур анализа и контроля, необходимые для проектирования алгоритмов.

Для решения задачи классификации необходимо указать тип объектов учета t_1 из множества типов V_t , которые необходимо разбить на классы. Также должно быть определено подмножество независимых показателей $X_{p1} = \{p_1, p_2, \dots, p_h, \dots, p_m\}$ данного типа объектов учета, на основании которых будет определяться значение зависимого показателя u , т.е. класс объекта учета. Наконец, должно быть определено множество значений, которые может принимать показатель u , тем самым определяя количество классов объектов учета.

Для построения функции классификации задается обучающая выборка данных. Для каждого значения класса вводится как минимум один набор

значений независимых показателей. Эти значения могут быть взяты как из базы данных мониторинга, так и введены вручную. Чем больше обучающая выборка, тем точнее будет функция классификации. Набор показателей X_{p1} объекта учета o_j рассматривается как координаты, а функция классификации имеет следующий вид [44]:

$$y_j = \omega_0 + \omega_1 p_1 + \omega_2 p_2 + \dots + \omega_m p_m,$$

где $\omega_0, \omega_1, \dots, \omega_m$ – веса независимых показателей, в поиске которых и состоит задача нахождения классификационной функции.

В результате вышеизложенного модель правила анализа для задачи классификации будет иметь вид следующей функции:

$$F(t, X_p, S, M) \rightarrow \Omega,$$

где на вход функции подаются:

- $t \in V_t$ – тип объектов учет среди которых идет классификация;
- X_p – набор независимых показателей, на основании которых будет определяться класс объекта учета;
- S – множество значений зависимого показателя;
- M – обучающая выборка, представляющая собой матрицу, в строках которой находятся значения независимых показателей и соответствующее значение зависимого показателя (класса);

результат функции $\Omega = \{ \omega_0, \omega_1, \dots, \omega_m \}$ – множество (вектор) весов независимых показателей.

При решении задачи кластеризации необходимо также, как и в задаче классификации, указать тип t_1 объектов учета, которые необходимо разбить на кластеры, и выбрать показатели $X_{p1} = \{ p_1, p_2, \dots, p_h, \dots, p_m \}$ данного типа, влияющие на процесс разбиения по кластерам. Необходимо сгруппировать данные таким образом, чтобы целевая функция алгоритма разбиения достигала экстремума (минимума). В модели правила кластеризации будут использоваться следующие понятия:

- входное множество данных M , на котором строится разбиение;
- метрика расстояния d ;

- вектор центров классов C (множество кластеров);
- матрица разбиения по кластерам U ;
- целевая функция $J=J(M,d,C,U)$

Исследуемые данные (входное множество данных) представляются как множество векторов в многомерном пространстве, описывающих объекты учета. Каждый вектор состоит из набора координат – показателей $X_{p1} = \{p_1, p_2, \dots, p_h, \dots, p_m\}$. Каждый показатель имеет свой тип (числовой или категориальный) и свое множество значений. Множество значений в случае числового показателя задается в виде одного или нескольких интервалов числовой оси. В случае категориального показателя множество значений задается перечислением возможных значений. Большое значение при подготовке данных к кластеризации имеет их очистка и нормировка. Очистка данных производится как по входному множеству в целом, так и по множеству показателей. Очистка данных по множеству показателей производится для исключения зависимых показателей, которые не окажут влияния на результат, но увеличат время, необходимое на обработку данных. Нормировка необходима для того, чтобы на результаты кластеризации не оказывали влияния различные области значений отдельных показателей. При этом числовые и категориальные показатели нормируются по-разному. В результате каждый показатель представляется значением из единичного интервала.

Метрика расстояния – важнейшее понятие, используемое в кластеризации. Именно с помощью расстояния между входными векторами показателей определяется сходство и различие объектов учета. Есть множество способов вычисления расстояния: евклидово, Хемминга, расстояние Махаланобиса и др. Выбор способа вычисления расстояния зависит от природы объектов учета и непосредственно влияет на результат.

Целевая функция – это функция, минимизация которой дает решение задачи кластеризации. Последовательность действий, реализующая поиск минимума целевой функции, является алгоритмом кластеризации.

Матрица разбиения – это основной результат кластеризации. Матрица разбиения представляет собой таблицу, где каждая ячейка содержит значение функции принадлежности данного вектора показателей заданному кластеру. На основании этой матрицы получается итоговое разбиение.

Помимо матрицы принадлежности в качестве результата порождается множество центров кластеров. Центр кластеров – это вектор, степень принадлежности которого заданному кластеру максимальна. Как правило, центров кластеров нет в исходном множестве.

В результате модель правила для задачи кластеризации будет иметь вид следующей функции [44]:

$$F(t, X_p, M, d) \rightarrow (C, U),$$

где на вход функции подаются:

– $t \in V_t$ – тип объектов учета которые разбиваются на кластеры;

– X_p – набор независимых показателей, на основании которых будет осуществляться разбиение;

– M – входное множество объектов учета;

– d – метрика расстояний;

на выходе – множество центров кластеров C и матрица разбиения по кластерам U .

Для решения задачи прогнозирования необходимо указать экземпляр объекта учета $o_1 \in V_o$ и выбрать периодический показатель p из множества показателей X_{p1} , сопоставленного с типом данного экземпляра объекта учета отношением $R_{p0} = (o_1, X_{p1})$. Для данного показателя задается множество его значений (количество периодов учета), на основании которых будет рассчитываться прогноз, и количество периодов, на которое необходимо рассчитать прогноз.

В результате модель правила для задачи прогнозирования будет иметь вид следующей функции [44]:

$$F(o, p, S_n, k) \rightarrow S_k,$$

где на вход функции подаются:

– $o_1 \in V_o$ – экземпляр объекта учета, имеющий показатель p для которого будут определяться прогнозные значения:

$S_n = \{s_1, s_2, \dots, s_i, \dots, s_n\}$ – множество значений периодического показателя p , на основании которых рассчитывается прогноз;

– k – количество периодов, на которое необходимо сделать прогноз;

на выходе функции $S_k = \{s_1, s_2, \dots, s_i, \dots, s_k\}$ – множество прогнозируемых значений периодического показателя.

2.4.3 Модели критериев контроля данных

Во-первых, отметим, что задача контроля значений показателей может решаться на следующих наборах исходных данных [44]:

– первичные показатели в базе данных;

– данные отчетов.

На логическом уровне модели критериев контроля для обоих источников данных будут одинаковые. Отличие в том, что в отчете присутствуют агрегированные (итоговые) показатели, что необходимо учитывать при задании параметров и величин, с которыми будут сравниваться показатели в процессе контроля данных.

Во-вторых, критерии контроля данных могут быть как четкими – дающими решение в рамках четкой логики, так и нечеткими – определяющими не факт, а степень принадлежности (близости) к некоторому решению задачи контроля данных. Модели контроля данных на основании четких и нечетких критериев будут существенно отличаться, поэтому их рассмотрение будет в отдельных нижеследующих подпунктах.

Модели четких критериев контроля данных.

Для контроля наименьшего и наибольшего значения показателей необходимо задать тип объектов учета $t_1 \in V_t$, и показатель, соотнесенный с данным типом объектов учета. Для этого показателя задаются максимальное и минимальное допустимые значения.

В результате модель критерия контроля максимального и минимального значений показателей будет иметь вид следующей функции [44]:

$$F(t, m, M, \max, \min) \rightarrow (P_m, P_n),$$

где на вход подаются:

– $t \in V_t$ – тип объекта учета

– m – показатель, принадлежащий типу объекта учета t ;

– M – множество объектов учета типа t , которые будут проходить контроль максимального и минимального значений по показателю m ;

– \max и \min – максимальное и минимальное допустимые значения для показателя m ;

на выходе имеются:

– P_m – множество объектов учета, которые удачно прошли контроль максимального и минимального значения по показателю m ;

– P_n – множество объектов учета, которые не прошли контроль максимального и минимального значения по показателю m .

Аналогично модель критерия контроля разницы между наибольшим и наименьшим значениями показателя будет иметь вид следующей функции [44]:

$$F(t, r, M, \text{int}) \rightarrow (P_r, P_n),$$

где на вход подаются:

– $t \in V_t$ – тип объекта учета;

– r – показатель, принадлежащий типу объекта учета t ;

– M – множество объектов учета типа t , которые будут проходить контроль разницы между наибольшим и наименьшим значениями для показателя r ;

– int – допустимая разница между наибольшим и наименьшим значениями показателя r ;

на выходе имеются P_r – множество объектов учета, которые удачно прошли контроль разницы между наибольшим и наименьшим значениями для показателя r ;

– P_n – множество объектов учета, которые не прошли контроль разницы между наибольшим и наименьшим значениями для показателя r .

Модель критерия контроля среднего арифметического будет иметь вид следующей функции [44]:

$$F(t, a, M, a_{\max}, a_{\min}) \rightarrow (P_a, P_n),$$

где на вход подаются:

– $t \in V_t$ – тип объекта учета;

– a – показатель, принадлежащий типу объекта учета t ;

– M – множество объектов учета типа t , которые будут проходить контроль среднего арифметического значений показателя a ;

– a_{\max} и a_{\min} – максимальное и минимальное допустимые значения среднего арифметического значений показателя a ;

на выходе имеются:

– P_a – множество объектов учета, которые удачно прошли контроль среднего арифметического значений показателя a ;

– P_n – множество объектов учета, которые не прошли контроль среднего арифметического значений показателя a .

Модель критерия контроля выбросов будет иметь вид следующей функции [44]:

$$F(t, g, M, out) \rightarrow (P_g, P_n),$$

где на вход подаются:

– $t \in V_t$ – тип объекта учета;

– g – показатель, принадлежащий типу объекта учета t ;

– M – множество объектов учета типа t , которые будут проходить контроль выбросов для показателя g ;

– out – допустимая разница между средним арифметическим значений показателя g и его новым значением;

на выходе имеются:

– P_g – множество объектов учета, которые удачно прошли контроль выбросов для показателя g ;

– P_n – множество объектов учета, которые не прошли контроль выбросов для показателя g .

Модели нечетких критериев контроля данных.

При решении задач контроля данных производится сравнение значений показателей с некоторыми запланированными или нормативными величинами. Другими словами значения показателей или их производные должны лежать в определенных диапазонах по различным критериям. В моделях четких критериев контроля данных объекты учета разбиваются на два множества. В одном множестве содержатся объекты учета, у которых значения показателей принадлежат заданному диапазону, в другом множестве содержатся объекты учета, у которых значения показателей не входят в заданный диапазон. Степень принадлежности значений показателя объекта учета заданному диапазону не учитывается.

В моделях нечетких критериев контроля данных вводится функция принадлежности следующего вида [44]:

$$N_k = \{ \langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle, \dots, \langle x_i, y_i \rangle, \dots, \langle x_k, y_k \rangle \},$$

где x_i – часть заданного диапазона (поддиапазон), а y_i – символьное имя, характеризующие объекты учета, у которых значения показателей попали в данную часть диапазона.

Необходимо отметить, что вид функции принадлежности может отличаться от приведенного выше в зависимости от специфики задачи контроля. Так, аналитика могут интересовать не значения, попавшие в диапазон корректных, а вышедшие за его пределы, и, соответственно, части 1..k будут выделяться на диапазоне не прошедших контроль значения, определяя степень отклонения. Возможна и комбинация этих двух видов функций, когда отслеживается близость к границам диапазона как с одной стороны, так и с другой.

В результате объекты учета разбиваются на $k+1$ множеств, в которых k множеств соответствуют принадлежности значений показателей объектов учета k частям заданного диапазона, и одно множество, в котором значения показателей объектов учета не принадлежат заданному диапазону.

В результате вышеизложенного модели нечетких критериев контроля данных будут иметь вид следующих функций (обозначения, за исключением P , идентичны приведенным в предыдущем подпункте):

– для задачи контроля наименьшего и наибольшего значений показателей:

$$F(t, m, M, \max, \min, N_k) \rightarrow P;$$

– для контроля разницы между наибольшим и наименьшим значениями показателей:

$$F(t, r, M, \text{int}, N_k) \rightarrow P;$$

– для задачи контроля среднего арифметического:

$$F(t, a, M, \text{amax}, \text{amin}, N_k) \rightarrow P;$$

– для задачи контроля выбросов – значений показателей, резко отличающихся от основного числа значений:

$$F(t, g, M, \text{out}, N_k) \rightarrow P;$$

где P – множество, состоящее из подмножеств, соответствующих выделенным в функции принадлежности k поддиапазнам, и одного или более подмножеств (зависит от постановки задачи контроля и вида диапазона, определенного N_k), соответствующих значениям, не вошедшим в диапазон.

2.5 Выводы по второй главе

1. В ходе анализа задач получения сводных данных, анализа и контроля данных была разработана модель процесса переработки данных, обеспечивающая возможность повышения абстракции представления объекта мониторинга и унификацию на этой основе представления исходных данных и результатов процесса переработки.

2. На основании модели процесса переработки данных была разработана методика построения процессов автоматизации анализа и контроля данных с возможностью выбора источника данных.

3. Для решения задач получения, анализа и контроля первичных и сводных данных системы административного мониторинга были разработаны формализованные модели правил анализа и критериев контроля данных, структуры и содержания выходных отчетных форм в терминах модели хранения данных системы административного мониторинга

4. Построение шаблона отчета должно быть реализовано (вследствие иерархической структуры объектов учета) в виде итеративной процедуры, обеспечивающей последовательную обработку объектов учета каждого из уровней.

3 Разработка и исследование алгоритмического и информационного обеспечения средств адаптации автоматизированной системы административного мониторинга

3.1 Разработка алгоритма формирования шаблона отчета

Учитывая нетривиальность получаемых конструкций поясним применение предложенного во второй главе языка описания шаблона отчетных форм, приведя алгоритм формирования шаблона отчета.

Задание шаблона ведется по уровням (поколениям) [45]. На каждом уровне определяются объекты учета одного поколения, их показатели, необходимость вести аналитический подсчет в случае, если показатели числовые или нет и прочее. Ниже приведено описание алгоритма задания нового шаблона и его графическое представление (рисунок 3.1).

Для каждого уровня иерархии объектов учета, начиная с корневого и заканчивая уровнем, для которого множество родителей, чьих потомков необходимо рассмотреть, будет пустым, выполняются следующие действия:

1 Отбор множества объектов учета для визуализации на текущем уровне. Для удобства пользователя и ускорения отбора возможны следующие укрупняющие варианты:

- показывать все объекты учета данного уровня,
- показывать объекты учета определенного типа (из типов, присутствующих на уровне);
- показать определенные экземпляры объектов учета.

Подобный интерфейс позволяет выбрать, например, объекты разного типа, если эти типы существуют в рассматриваемом поколении.

Указание объекта учета может происходить как без фильтрации (простой выбор из списка), так и с помощью фильтра.

Обозначим полученное множество объектов учета O_{i_render} , где i – номер поколения.

2 Отбор показателей для O_{i_render} , по которым необходимо отображать значения на данном уровне. На этом этапе определяется, будет ли отчет простым (без обработки) или сложным. Ведение обработки значений показателей и вычисление агрегирующих функций возможны только на числовых показателях. Поэтому возможны следующие варианты включения значений показателей на данном уровне:

- использовать числовые показатели (с возможностью агрегации, как на данном уровне, так и на всем пути от корня дерева до рассматриваемого уровня);

- использовать исходные значения показателей (только вывод значения, без возможности агрегации и другой обработки);

- не использовать значения показателей объектов учета в отчете на данном уровне (для отчетов, требующих только выявления родственных связей объектов учета).

Как и в случае с объектами учета, указание показателей может происходить как без фильтрации (простой выбор из списка), так и с использованием фильтра.

3 Если выбраны числовые показатели, то указывается метод обработки данных на текущем уровне. На данном этапе при доступности только данных объектов учета одного поколения, группировка для проведения агрегации возможна по следующим наборам объектов учета:

- группировка данных по всем объектам учета на данном уровне;

- группировка данных по типам (агрегирующие и другие функции обработки выдают значения для каждой группы однотипных объектов учета данного уровня);

- группировка данных объектов учета, имеющих одного родителя.

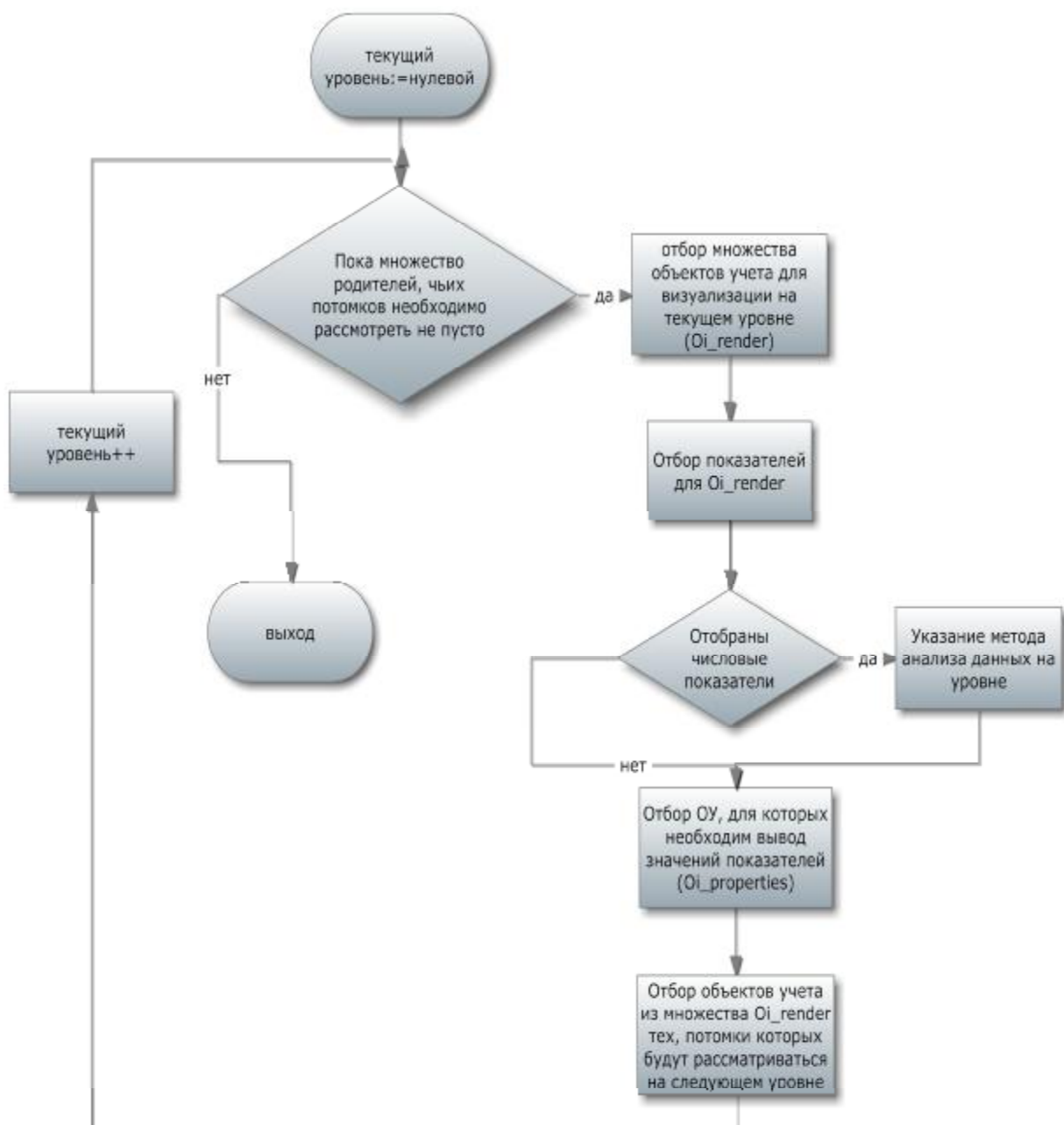


Рисунок 3.1 – Алгоритм формирования шаблона отчета

После определения множества обрабатываемых данных возможен непосредственный выбор функции обработки, например, вычисление суммы значений показателей, среднего значения и др.

4 Из выбранного множества объектов учета O_{i_render} , присутствующих в отчете на текущем уровне, осуществляется отбор тех, для которых требуется осуществлять вывод значений указанных выше показателей. Для остальных

объектов учета будут отображаться только названия. Обозначим данное множество объектов учета $O_{i_properties}$ ($O_{i_properties} \subset O_{i_render}$).

5 Подготовка для перехода на следующий уровень. Для выполнения алгоритма на следующем уровне иерархии объектов учета необходимо сформировать множество объектов учета текущего уровня, потомки которых подлежат рассмотрению на следующем. Обозначим данное множество $O_{i_navigate}$ ($O_{i_navigate} \subset O_{i_render}$).

Для удобства пользователя помимо конструктора шаблонов целесообразно предусмотреть также типовые отчеты, которые представляют собой конкретные указания по наполнению данными и визуализации заранее заданной структуры шаблонов. Для их использования достаточно указать необходимые параметры: типы объектов учета, объекты учета и показатели (в случае каждого конкретного предустановленного шаблона набор задаваемых параметров может отличаться). Задавать структуру шаблона в этом случае уже нет необходимости.

3.2 Методы анализа данных

Методы прогнозирования будут рассмотрены в следующем пункте, из-за особенностей их применения к задачам административного мониторинга. При настройке правил анализа данных для задачи кластеризации и классификации задаются все необходимые параметры для методов решения этих задач. Эти параметры не требуют специальных знаний самих методов, а характеризуют лишь саму задачу. В отличие от методов задачи классификации и кластеризации в методах задачи прогнозирования, необходимы специальные знания самих методов для настройки их параметров. Из-за возможности системы административного мониторинга адаптироваться под изменения требований задачи мониторинга, заранее нельзя предположить какой метод прогнозирования использовать для прогнозирования состояния объекта мониторинга.

3.2.1 Метод наименьших квадратов при решении задачи классификации

При построении математической функции классификации или регрессии основная задача сводится к выбору наилучшей функции из всего множества вариантов. Дело в том, что может существовать множество функций, одинаково классифицирующих одну и ту же обучающую выборку.

В результате задачу построения функции классификации и регрессии можно формально описать как задачу выбора функции с минимальной степенью ошибки [31]:

$$\min_{f \in F} R(f) = \frac{1}{m} \sum_{i=1}^m c(y_i, f(p_i)), \quad (6.1)$$

Где F – множество всех возможных функций; $c(y_i, f(p_i))$ – функция потерь, в которой $f(p_i)$ – значение зависимого показателя, найденное с помощью функции f для вектора $p_i \in M$, а y_i – ее точное (известное) значение.

При оценке величин, принимающих вещественные значения, целесообразно использовать разность $f(p)$ - y для оценки качества классификации. Чаще всего применяется минимизация квадратов разностей $f(p)$ - y . Соответствующим образом минимизируем:

$$c(y, f(p)) = (f(p) - y)^2. \quad (6.2)$$

При решении задачи классификации коэффициенты ω можно вычислить, используя квадратичную функцию потерь (6.2) и множество линейных функций F :

$$F := \left\{ f \mid f(p) = \sum_{i=1}^n w_i f_i(p), w_i \in \mathfrak{R} \right\},$$

где $f_i: P \rightarrow R$.

Необходимо найти решение следующей задачи:

$$\min_{f \in F} R(f) = \min_{w \in \mathfrak{R}^n} \frac{1}{m} \sum_{i=1}^m \left(y_i - \sum_{j=1}^n w_j f_j(p_i) \right)^2,$$

где n – количество независимых показателей.

Вычисляя производную $R(f)$ по ω и вводя обозначение $M_{ij}:=f_i(p_j)$, получаем, что минимум достигим при условии:

$$M^T y = M^T M w.$$

Решением этого выражения будет[46]:

$$w = (M^T M)^{-1} M^T y.$$

Очевидно, что все показатели должны быть представлены в виде числовых параметров. Для преобразования логических и категориальных показателей к числовым используют разные способы. Логические, как правило, кодируют цифрами 1 и 0. Истине ставят в соответствие 1, а ложь обозначают 0. Значениями категориальных показателей являются имена возможных состояний объекта учета по определенному критерию. Разумеется, таких состояний может быть больше двух. Их имена должны быть перечислены и пронумерованы в списке. Каждое имя из списка должно быть представлено своим номером. В итоге категориальный показатель преобразуется в числовой.

3.2.2 Решение задачи кластеризации с использованием метода k-средних

Вначале выбираются k произвольных исходных центров кластеров среди точек в пространстве (векторов показателей) всех объектов учета. Не очень критично, какие именно это будут центры, процедура выбора исходных точек отразится, главным образом, только на времени счета. Например, это могут быть первые k объектов входного множества M . Далее итерационно выполняются операции двух шагов [31].

На первом шаге все объекты разбиваются на k групп, наиболее близких к одному из центров. Близость определяется расстоянием, которое вычисляется одним из следующих способов.

Евклидово расстояние:

$$d_2(x_i, x_j) = \sqrt{\sum_{i=1}^m (x_{ij} - x_{ji})^2}.$$

Расстояние по Хеммингу. Это расстояние является просто средним разностей по координатам. В большинстве случаев данная мера расстояния приводит к таким же результатам, как и для обычного расстояния Евклида, однако для нее влияние отдельных больших разностей (выбросов) уменьшается (так как они не возводятся в квадрат). Расстояние по Хеммингу вычисляется по следующей формуле:

$$d_H(x_i, x_j) = \sum_{i=1}^m |x_{ij} - x_{ji}|.$$

Расстояние Чебышева. Это расстояние может оказаться полезным, когда желают определить два объекта как «различные», если они различаются по какой-либо одной координате (какому-либо показателю). Расстояние Чебышева вычисляется по формуле:

$$d_\infty(x_i, x_j) = \max_{1 \leq t \leq m} |x_{it} - x_{jt}|.$$

Пиковое расстояние предполагает независимость между случайными переменными, что говорит о расстоянии в ортогональном пространстве. Но в практических приложениях эти переменные не являются независимыми:

$$d_L(x_i, x_j) = \frac{1}{m} \cdot \sum_{i=1}^m \frac{|x_{it} - x_{jt}|}{x_{it} + x_{jt}}.$$

На втором шаге вычисляются новые центры кластеров. Центры можно вычислить как средние значения показателей объектов учета, отнесенных к сформированным группам. Новые центры, естественно, могут отличаться от предыдущих.

Рассмотренная операция повторяется итеративно до тех пор, пока центры кластеров (соответственно, и границы между ними) не перестанут меняться. Далее приводится формальное описание алгоритма k-средних.

Базовые определения и понятия в рамках данного алгоритма имеют вид:

- множество входных объектов $M = \{m_j\}_{j=1}^d$, d – количество входных объектов учета (векторов показателей);
- метрика расстояния, расстояния рассчитываемая по формуле:

$$d_A^2(m_j, c^{(i)}) = \|m_j - c^{(i)}\|_A^2,$$

где матрица A определяет способ вычисления расстояния;

– вектор центров кластеров $C = \{c^{(i)}\}_{i=1}^c$, где:

$$c^{(i)} = \frac{\sum_{j=1}^d u_{ij} m_j}{\sum_{j=1}^d u_{ij}}, \quad 1 \leq i \leq c;$$

– матрица разбиения по кластерам $U = \{u_{ij}\}$, где:

$$u_{ij} = \begin{cases} 1, & d(m_j, c^{(i)}) = \min_{1 \leq k \leq c} d(m_j, c^k), \\ 0, & \text{иначе;} \end{cases}$$

– целевая функция:

$$J(M, U, C) = \sum_{i=1}^c \sum_{j=1}^d u_{ij} d_A^2(m_j, c^{(i)});$$

– набор ограничений:

$$u_{ij} \in \{0, 1\}; \quad \sum_{i=1}^c u_{ij} = 1; \quad 0 < \sum_{j=1}^d u_{ij} < d,$$

который определяет, что каждый вектор показателей может принадлежать только одному кластеру и не принадлежать остальным. В каждом кластере содержится не менее одного объекта учета, но менее общего количества объектов учета.

Конструктивно алгоритм представляет собой итерационную процедуру следующего вида.

1. Проинициализировать начальное разбиение (например, случайным образом), выбрать точность δ (используется в условии завершения алгоритма), проинициализировать номер итерации $l=0$.

2. Определить центры кластеров по следующей формуле:

$$c_l^{(i)} = \frac{\sum_{j=1}^d u_{ij}^{(l-1)} m_j}{\sum_{j=1}^d u_{ij}^{(l-1)}}, \quad 1 \leq i \leq c.$$

3. Обновить матрицу разбиения с тем, чтобы минимизировать квадраты ошибок, используя формулу:

$$u_{ij}^{(l)} = \begin{cases} 1, d(m_j, c_i^{(l)}) = \min_{1 \leq k \leq c} d(m_j, c_k^{(l)}), \\ 0, \text{иначе.} \end{cases}$$

4. Проверить условие $\|U^{(l)} - U^{(l-1)}\| < d$. Если условие выполняется, завершить процесс, если нет – перейти к шагу 2 с номером итерации $l=l+1$.

3.3 Алгоритмы контроля данных

3.3.1 Алгоритм контроля наименьшего и наибольшего значений показателей

Алгоритм контроля наименьшего и наибольшего значений показателей будет иметь следующий вид:

1. Получить M множество объектов учета типа t .
2. Выбрать объект учета из множества M .
3. Для выбранного объекта учета сравнить текущее значение показателя m с минимальным и максимальным допустимыми значениями для данного показателя (\max и \min).
4. Если значение показателя $m > \max$ или $m < \min$, то поместить объект учета во множество P_n . Иначе поместить объект учета во множество P_m .
5. Если множество M непустое перейти к шагу 2.
6. Вывести список объектов учета из множества P_n , которые неудачно прошли контроль наименьшего и наибольшего значений показателей. Вывести список объектов учета из множества P_m , которые удачно прошли контроль наименьшего и наибольшего значений показателей.

3.3.2 Алгоритм контроля разницы между наибольшим и наименьшим значениями показателей

Алгоритм контроля разницы между наибольшим и наименьшим значениями показателей имеет следующий вид:

1. Получить M множество объектов учета типа t .
2. Выбрать объект учета из множества M .
3. Для показателя g выбранного объекта учета найти максимальное и минимальное значения (s_{\max} и s_{\min}).
4. Сравнить модуль разности найденных значений с допустимой разницей между наибольшим и наименьшим значениями для показателя g .
5. Если $|s_{\max} - s_{\min}| > \text{int}$, то поместить объект учета во множество P_n .
Иначе поместить объект учета во множество P_r .
6. Если множество M непустое перейти к шагу 2.
7. Вывести список объектов учета из множества P_n , которые неудачно прошли контроль разницы между наибольшим и наименьшим значениями для показателя g . Вывести список объектов учета из множества P_r , которые удачно прошли контроль разницы между наибольшим и наименьшим значениями для показателя g .

3.3.3 Алгоритм контроля среднего арифметического

Алгоритм контроля среднего арифметического выглядит следующим образом:

1. Получить M множество объектов учета типа t .
2. Выбрать объект учета из множества M .
3. Для показателя a выбранного объекта учета найти среднее арифметическое его значений.
4. Сравнить найденное среднее арифметическое с минимальным и максимальным допустимыми значениями среднего арифметического для показателя a (a_{\min} и a_{\max}).
5. Если $\text{avr}(k, S) > a_{\max}$ или $\text{avr}(k, S) < a_{\min}$, то поместить объект учета во множество P_n . Иначе поместить объект учета во множество P_a .
6. Если множество M непустое перейти к шагу 2.
7. Вывести список объектов учета из множества P_n , которые неудачно прошли контроль среднего арифметического значений показателя a . Вывести

список объектов учета из множества P_a , которые удачно прошли контроль среднего арифметического значений показателя a .

3.3.4 Алгоритм контроля выбросов

Алгоритм контроля выбросов – значений показателей, резко отличающихся от основного числа значений, выглядит следующим образом:

1. Получить M множество объектов учета типа t .
2. Выбрать объект учета из множества M .
3. Для показателя g выбранного объекта учета найти среднее арифметическое его значений за исключением текущего.
4. Сравнить разность между найденным средним арифметическим и текущим значением показателя g с допустимым значением выбросов для показателя g .
5. Если $|\text{avg}(h, S) - s_{h+1}| > \text{out}$, то поместить объект учета во множество P_n . Иначе поместить объект учета во множество P_g .
6. Если множество M непустое перейти к шагу 2.
7. Вывести список объектов учета из множества P_n , которые неудачно прошли контроль выбросов для показателя g . Вывести список объектов учета из множества P_g , которые удачно прошли контроль выбросов для показателя g .

При использовании моделей нечетких критериев контроля данных отличие их алгоритмов от вышеописанных будет заключаться в том, что для объектов учета будет вычисляться значение функции принадлежности, которая разделит объекты учета на заданные функцией подмножества.

3.4 Автоматизация настройки задачи прогнозирования состояния объекта мониторинга

В системе мониторинга выходные наборы данных представляют собой множество показателей объектов учета, значения которых фиксируются

через равные промежутки времени – периоды. Поэтому выходные наборы данных мониторинга можно рассматривать как временные ряды.

Для анализа временных рядов чаще всего применяется аппарат математической статистики, что приводит к необходимости наличия у специалиста или исследователя знаний в области соответствующих методов и моделей прогнозирования, поскольку использование нерациональной модели может дать правдоподобный, но далекий от реальности результат. Это обусловлено тем, что выполнение численных операций в различных моделях не приводит к алгебраическим ошибкам и, следовательно, применение ошибочной модели зачастую незаметно. Использование статистических пакетов так же не позволяет избежать описанной проблемы, поскольку в них автоматизировано выполняются лишь вычисление по выбранной модели (алгоритму) и формальная проверка входных данных. Кроме того данные программные продукты обычно обладают сложным пользовательским интерфейсом, возможностью настройки большого количества параметров мало знакомых обычному пользователю [47].

Для прогнозирования временных рядов разработано большое количество методов или алгоритмов. Результат работы каждого алгоритма зависит от свойств прогнозируемого ряда, поэтому возникает задача автоматического или в существенной степени автоматизированного выбора наилучшего алгоритма из некоторого заданного семейства.

3.4.1 Задача построения классификационной функции для выбора метода прогнозирования

Описанную выше задачу можно рассматривать как задачу классификации. Объектами классификации являются временные ряды. Для метаописания временного ряда создается набор признаков, характеризующий временной ряд. Метками классов являются название алгоритмов прогнозирования. Временной ряд относится к некоторому классу, если соответствующий этому классу алгоритм работает на временном ряде

наилучшим образом по заданному функционалу качества. Решение такой задачи классификации описывается в [48]. Для классификации был использован алгоритм k взвешенных ближайших соседей. В качестве признаков были использованы длина ряда, число вспомогательных рядов, максимальное и минимальное значения ряда, число пропущенных значений, среднее значение ряда. Автором был проведен вычислительный эксперимент на выборке из 120 рядов, по результатам прогнозирования их тремя алгоритмами – SSA, ARIMA, локальными методами прогнозирования. Исходя из результатов эксперимента, делается вывод, что для более успешного автоматического выбора алгоритмов прогнозирования «требуется более тщательный выбор признаков для метаописания рядов, а также, возможно, следует использовать другие алгоритмы классификации» [48].

Чтобы произвести «тщательный» выбор признаков для метаописания временных рядов необходимо расширить набор признаков. В качестве признаков могут выступать «стандартные» характеристики выходных наборов данных мониторинга [49], а также, в случае анализа данных административного мониторинга, характеристики структуры модели хранения системы административного мониторинга, разработанной в Госуниверситете-УНПК в результате исследования подходов к созданию адаптируемого программного обеспечения систем сбора, хранения и обработки данных. Логическая модель хранения данных административного мониторинга специфична [25], отражает особенности конкретной задачи административного мониторинга, но при этом реализуется на физическом уровне универсальными средствами. Следовательно, можно использовать характеристики ее структуры в качестве признаков классификации.

Также для успешной классификации необходимо определить оптимальное количество классов, т.е. семейство самых распространенных методов прогнозирования, которые по возможности компенсировали недостатки друг друга.

После того как будет получен полный список признаков классификации для решения задачи автоматического выбора методов (алгоритмов) прогнозирования из заданного семейства, необходимо осуществить поиск оптимального набора признаков на всем пространстве признаков. В результате получится множество классификационных функций, среди которых нужно отобрать функции с приемлемым процентом ошибок. Для автоматической настройки параметров методов прогнозирования на конкретный временной ряд необходимо установить зависимость между признаками, принадлежащими отобранным классификационным функциям, и параметрами прогнозной модели каждого метода (алгоритма) прогнозирования.

Ниже приводится формальная постановка задачи построения классификационной функции для выбора метода прогнозирования [50].

Пусть задано множество алгоритмов прогнозирования $A = \{a_1, a_2, \dots, a_i, \dots, a_n\}$, задано множество классификационных признаков $X = \{x_1, x_2, \dots, x_j, \dots, x_m\}$, задано множество временных рядов $T = \{t_1, t_2, \dots, t_k, \dots, t_h\}$. Для каждого временного ряда задано его признаковое описание $t_i = \{t_{i1}, t_{i2}, \dots, t_{ij}, \dots, t_{im}\}$ и алгоритм прогнозирования $t_{iopt} \hat{I} A$, работающий наилучшим образом на данном временном ряде по заданному функционалу качества. Необходимо найти множество весов для классификационных признаков $W = \{w_0, w_1, \dots, w_j, \dots, w_m\}$ такое, что:

$$\frac{r}{h} \rightarrow \max,$$

где r – мощность множества правильно классифицированных рядов $Y \hat{I} T$, $t_i \hat{I} Y$, если $t_{iopt} = w_0 + w_1 t_{i1} + w_2 t_{i2} + \dots + w_m t_{im}$; h – мощность множества временных рядов.

Множество T является обучающей выборкой. Для того чтобы T состояло из временных рядов, содержащих любое сочетание m классификационных признаков, тем самым образуя полный факторный

план с двумя значениями уровней факторов (максимальное и минимальное значения признаков), его мощность должна быть равна $h=2^m$.

3.4.2 Исследование характеристик выходных наборов данных системы административного мониторинга

Выше было сказано, что выходные наборы данных мониторинга можно рассматривать как временные ряды, а также что в качестве признаков могут выступать «стандартные» характеристики временных рядов, а, в случае анализа данных административного мониторинга, характеристики структуры модели хранения системы административного мониторинга.

3.4.2.1 Характеристики временного ряда, обуславливающие выбор прогнозной модели и ее параметров

Ниже перечислены характеристики временного ряда, которые могут каким-либо образом повлиять на выбор прогнозной модели [49]:

1. Наличие или отсутствие тренда.
2. Наличие или отсутствие сезонной компоненты.
3. Наличие или отсутствие случайной компоненты.
4. Длина ряда. Под длиной временного ряда понимают количество входящих в него уровней. Обычно прогнозирование временных рядов дает качественный результат, если длина ряда достаточна велика.
5. Количество аномальных наблюдений (выбросов). Наличие аномальных наблюдений приводит к искажению результатов моделирования временных рядов. Для диагностики аномальных наблюдений разработаны различные критерии, например, метод Ирвина.
6. Число пропущенных значений. Существуют специальные методы прогнозирования временных рядов с пропущенными значениями (метод «гусеница») [51].
7. Число вспомогательных рядов [48].
8. Максимальное и минимальное значение временного ряда.

9. Среднее значение временного ряда.

10. Характеристики динамики изменения уровней временного ряда: средний абсолютный прирост, средний темп роста, средний темп прироста.

11. Стандартное отклонение.

3.4.2.2 Исследование структур хранения данных в системе административного мониторинга

Анализ различных практических задач административного мониторинга и соответствующих им моделей хранения данных показывает, что по критерию влияния структуры хранения на характеристики ряда значений показателей можно выделить следующие классы структур хранения данных [52]:

1. Терминальный (изолированный) объект учета, т.е. объект учета не имеющий дочерних объектов учета (ОУ).

2. Поддерево объектов учета (или подсеть с одним истоком и одним или несколькими стоками), в котором исследуемый объект учета является корнем (истоком) и не связан с потомками (не обязательно прямыми) по рассматриваемому показателю.

3. Поддерево объектов учета (или подсеть с одним истоком и одним или несколькими стоками), в котором исследуемый объект учета является корнем (истоком) и имеет связь с потомками (не обязательно прямыми) по рассматриваемому показателю.

Ниже дано формализованное определение двух последних классов.

Обозначим через V_{tree} непустое ($|V_{tree}| = k > 0$) множество потомков объекта учета o_0 . Пусть для o_0 задано отношение ассоциации $r_{p0} = (o_0, X_{p0})$. С объектом o_0 ассоциирован показатель $p \in X_{p0}$ – исследуемый показатель. Для всех $o_i \in V_{tree}$ ($i = 1..k$) заданы отношения ассоциации $r_{poi} = (o_i, X_{pi})$, где $X_{pi} \subset P$ – подмножество показателей, ассоциированных с объектом учета o_i .

Если $\neg \exists o_i \in V_{tree}, \{p\} \cap X_{pi} = \{p\}$, то o_0 не имеет потомков, связанных с ним по показателю p_0 , и структура относится ко второму классу.

Если $\exists o_i \in V_{tree}, \{p\} \cap X_{pi} = \{p\}$, то o_0 имеет потомков, связанных с ним по показателю p_0 , и структура относится к третьему классу.

3.4.2.3 Выявления и описание «значащих» характеристик структуры модели хранения

Отношение модели хранения к одному из приведенных выше классов является качественным показателем, не в полной мере отражающим влияние конкретной модели объекта мониторинга (при рассмотрении конкретного объекта учета и конкретного показателя).

Поэтому для идентификации класса структуры хранения и для учета специфических свойств конкретной модели объекта мониторинга введем следующие количественные характеристики[52]:

1. Количество терминальных объектов учета N . Более точно – количество объектов учета – листьев, агрегированных в объекте учета, для которого осуществляется прогноз.

2. Количество связей с терминальными объектами учета по общему показателю L – это сумма длин путей от исследуемого объекта учета до каждого терминального объекта учета, связанного с первым по рассматриваемому показателю.

Перечисленные количественные характеристики позволяют однозначно идентифицировать класс структуры хранения:

– если ОУ не имеет дочерних ОУ, то есть $N=0$, то структура хранения данных для данной конкретной задачи представляется множеством изолированных ОУ;

– если у ОУ терминальные ОУ ($N>1$), но нет связей с терминальными ОУ по общему показателю ($L=0$), то структура хранения данных является сетью (деревом) объектов, не связанных по показателям;

– если у ОУ есть связи с терминальными ОУ по общему показателю ($N \geq 0$ и $L > 0$), то структура хранения данных является сетью (деревом) объектов, связанных по показателям.

Характеристики N и L позволяют не только классифицировать структуру хранения, но и получить дополнительную информацию о ней, позиционируя ее в широком пространстве от линейного списка объектов учета до некоторой сильносвязной сети.

Также (основываясь на результатах анализа различных наборов данных, полученных в ходе решения различных задач административного мониторинга) для учета специфических свойств конкретной модели объекта мониторинга введем дополнительные количественные характеристики:

3. Уровень объекта учета в дереве (сети) объектов учета.

4. Период учета показателя. Каждый динамический показатель в системе административного мониторинга характеризуется периодичностью, с которой осуществляется изменение его значения.

Все четыре характеристики структуры модели хранения данных могут использоваться в качестве расширенного набора признаков классификации. Могут быть отмечены следующие (формулируемые на данном этапе исследования в качественном виде) закономерности влияния характеристик структуры хранения данных об объекте мониторинга[52]:

1. Чем выше уровень объекта учета, чем больше количество терминальных объектов учета для рассматриваемого ОУ, чем больше количество связей объекта учета с терминальными ОУ по общему показателю (причем каждый из выделенных факторов может оказывать самостоятельное влияние на характеристики ряда), тем вероятнее отсутствие выбросов (аномальных наблюдений), тем меньше вклад недетерминированной составляющей во временной ряд, представленный показателем, характеризующим рассматриваемый ОУ.

2. Если период учета равен году и более, то можно говорить об отсутствии сезонной компоненты.

3.4.3 Методы прогнозирования временных рядов

Метод прогнозирования содержит последовательность действий, в результате выполнения которой определяется модель прогнозирования конкретного временного ряда.

Моделью прогнозирования временного ряда является функциональное представление, адекватно описывающее временной ряд.

В настоящее время насчитывается свыше 100 классов моделей. Число общих классов моделей, которые в тех или иных вариациях повторяются в других, гораздо меньше. Часть моделей и соответствующих методов относится к отдельным процедурам прогнозирования. Часть методов представляет набор отдельных приемов, отличающихся от базовых или друг от друга количеством частных приемов и последовательностью их применения [53].

В аналитическом обзоре [54] все методы прогнозирования делятся на две группы: интуитивные и формализованные.

Интуитивное прогнозирование применяется тогда, когда объект прогнозирования либо слишком прост, либо, напротив, настолько сложен, что аналитически учесть влияние внешних факторов невозможно. Интуитивные методы прогнозирования не предполагают разработку моделей прогнозирования и отражают индивидуальные суждения специалистов (экспертов) относительно перспектив развития процесса. Интуитивные методы основаны на мобилизации профессионального опыта и интуиции. Такие методы используются для анализа процессов, развитие которых либо полностью, либо частично не поддается математической формализации, то есть для которых трудно разработать адекватную модель. К таким методам относятся методы экспертных оценок, исторических аналогий, предвидения по образцу. Кроме того, в настоящее время широко распространено применение экспертных систем, в том числе с использованием нечеткой логики [53].

Этот класс методов нельзя применять «автоматически», без анализа лицом, принимающим решения, специфики наборов данных, поэтому он в дальнейшем рассматриваться не будет.

Формализованные методы рассматривают модели прогнозирования. В обзоре [55] модели прогнозирования разделяются на статистические модели и структурные модели.

В статистических моделях функциональная зависимость между будущими и фактическими значениями временного ряда, а также внешними факторами задана аналитически. К статистическим моделям относятся следующие группы:

- регрессионные модели;
- авторегрессионные модели;
- модели экспоненциального сглаживания.

В структурных моделях функциональная зависимость между будущими и фактическими значениями временного ряда, а также внешними факторами задана структурно. К структурным моделям относятся следующие группы:

- нейросетевые модели;
- модели на базе цепей Маркова;
- модели на базе классификационно-регрессионных деревьев.

Также к формализованным методам прогнозирования можно отнести метод «Гусеница»-SSA. Этот подход зародился в 70х-80х годах прошлого столетия. В его основе лежит трансформация ряда в матрицу и ее сингулярное разложение. После идентификации компонент сингулярного разложения происходит их группировка, приводящая к разложению исходного ряда на аддитивные компоненты, такие как тренд, колебания (периодики) и шум. В зарубежной литературе метод наиболее известен под названием SSA (Singular Spectrum Analysis), он возник из теории динамических систем [56].

Кроме того, необходимо отметить, что для узкоспециализированных задач иногда применяются особые модели прогнозирования. Так, например, для задачи прогнозирования уровня сахара крови человека применяются модели на основе дифференциальных уравнений. Для задачи прогнозирования транспортного потока, которая в последние несколько лет актуальна для мегаполисов, применяются гидродинамические модели. Для прогнозирования природных явлений, таких как землетрясения, применяется, например, модель, в основу которой положены нелинейные клетки (или соты), находящиеся под воздействием внешнего поля, и у которых есть внутреннее состояние, изменяющееся во времени под воздействием этого поля. Аналогичные модели разрабатываются и применяются для специальных процессов и систем. В рамках настоящей работы данный класс формализованных моделей не рассматривается [53].

Этот класс особых узкоспециализированных моделей прогнозирования, не подходит для решения задач прогнозирования по результатам административного мониторинга, так как их реализация будет слишком трудозатратна относительно их частоты использования.

Исходя из анализа характерных свойств наборов данных административного мониторинга и приведенного выше обзора формализованные методы пригодны для прогнозирования изменения состояния объекта мониторинга или его отдельных элементов (с точки зрения специфики наборов данных), так как они рассматривают модели (количественного) прогнозирования.

3.4.3.1 Выбор моделей прогнозирования с точки зрения их «автоматического» применения

Рассмотрим формализованные методы прогнозирования более подробно.

Линейная регрессионная модель. Самым простым вариантом регрессионной модели является линейная регрессия. В основу модели

положено предположение, что существует дискретный внешний фактор $X(t)$, оказывающий влияние на исследуемый процесс $Z(t)$, при этом связь между процессом и внешним фактором линейна. Для получения прогнозных значений $Z(t)$ в момент времени t необходимо иметь значение $X(t)$ в тот же момент времени t , что редко выполнимо на практике.

В основу нелинейной регрессионной модели положено предположение о том, что существует известная функция, описывающая зависимость между исходным процессом $Z(t)$ и внешним фактором $X(t)$. Однако на практике редко встречаются процессы, для которых вид функциональной зависимости между процессом $Z(t)$ и внешним фактором $X(t)$ заранее известен. В связи с этим нелинейные регрессионные модели применяются редко [53].

В базовом методе «Гусеница»-SSA есть два параметра. Первый – это целое число длина гусеницы, а второй параметр является структурным – это способ группировки главных компонент. Для анализа временного ряда не имеет смысла брать длину гусеницы, большую чем половина длины ряда. Чем больше длина гусеницы тем больше детальное разложение исходного ряда. Таким образом, наиболее детальное разложение достигается при выборе длины гусеницы, приблизительно равной половине длины ряда. Маленькая длина гусеницы может привести к смешиванию интерпретируемых компонент ряда. При решении задачи выделения периодической компоненты следует выбирать длину гусеницы кратной периоду [57].

Что касается выбора второго параметра, то процедуру разделения членов сингулярного разложения на группы (шаг группировки) трудно формализовать. Эта процедура основана на анализе сингулярных векторов и собственных чисел в сингулярных разложениях [58]. Поэтому с точки зрения автоматического применения метода «Гусеница»-SSA этот параметр должен находиться перебором.

В основу экспоненциального сглаживания заложена идея постоянного пересмотра прогнозных значений по мере поступления фактических. Модель экспоненциального сглаживания присваивает экспоненциально убывающие

веса наблюдениям по мере их старения. Таким образом, последние доступные наблюдения имеют большее влияние на прогнозное значение, чем старшие наблюдения. Модель Хольта-Винтерса или тройное экспоненциальное сглаживание применяется для процессов, которые имеют тренд и сезонную составляющую [53]. Эта модель из семейства моделей экспоненциального сглаживания больше всего подходит для прогнозирования изменения состояния объекта мониторинга, так учитывает наличие тренда и сезонной составляющей во временном ряде. На вход модель получает четыре параметра: α, β, γ – постоянные сглаживания, α – коэффициент сглаживания уровня, β – коэффициент сглаживания тренда, γ – коэффициент сглаживания сезонной составляющей, и четвертый параметр определяется длиной сезона исследуемого процесса.

В простом экспоненциальном сглаживании (метод Хольта) если α равно 1, то предыдущие наблюдения полностью игнорируются. Если α равно 0, то игнорируются текущие наблюдения. Значения α между 0 и 1 дают промежуточный результат [31].

Постоянные сглаживания в методе Хольта-Винтерса играют ту же роль, что и постоянная в простом экспоненциальном сглаживании. Подбираются они, например, путем перебора по этим параметрам с определенным шагом [59].

В отличие от методик прогнозирования временных рядов, основанных на экспоненциальном сглаживании, в методологии ARIMA не предполагается какой-либо четкой модели для прогнозирования данной временной серии. Задается лишь общий класс моделей, описывающих временной ряд и позволяющих выражать текущее значение переменной через ее предыдущие значения [59]. Затем алгоритм, подстраивая внутренние параметры, сам выбирает наиболее подходящую модель прогнозирования. Общая модель, предложенная Боксом и Дженкинсом, включает как параметры авторегрессии, так и параметры скользящего среднего [60]. Таким образом, имеется три типа параметров модели: параметры авторегрессии p ,

порядок разности r , параметры скользящего среднего q . В обозначениях Бокса и Дженкинса модель записывается как ARIMA (p, q, r).

Для модели ARIMA необходимо, чтобы ряд был стационарным, это означает, что его среднее постоянно, а выборочные дисперсия и автокорреляция не меняются во времени. Поэтому обычно необходимо брать разности ряда до тех пор, пока он не станет стационарным. Число разностей, которые были взяты, чтобы достичь стационарности, определяются параметром r . Для того чтобы определить необходимый порядок разности, нужно исследовать график ряда и автокоррелограмму. Сильные изменения уровня обычно требуют взятия несезонной разности первого порядка. Сильные изменения наклона требуют взятия разности второго порядка. Сезонная составляющая требует взятия соответствующей сезонной разности. Если имеется медленное убывание выборочных коэффициентов автокорреляции в зависимости от лага, обычно берут разность первого порядка [59]. На практике r необходимо определять, однако в большинстве случаев $r \leq 2$ [53].

На этапе идентификации порядка модели также необходимо решить, как много параметров авторегрессии (p) и скользящего среднего (q) должно присутствовать в эффективной модели процесса. На практике очень редко бывает, что число параметров p или q больше двух [59].

С точки зрения автоматического применения модели ARIMA для прогнозирования изменения состояния объекта мониторинга необходимо перебрать значения трех параметров в p, q, r [61].

В настоящее время самой популярной среди структурных моделей является модель на основе искусственных нейронных сетей [53]. Нейронные сети состоят из нейронов. Способ связи нейронов определяет архитектуру нейронной сети. В зависимости от способа связи нейронов сети делятся на:

- однослойные нейронные сети прямого распространения;
- многослойные нейронные сети прямого распространения;
- рекуррентные нейронные сети.

При помощи нейронных сетей возможно моделирование нелинейной зависимости будущего значения временного ряда от его фактических значений и от значений внешних факторов. Нелинейная зависимость определяется структурой сети и функцией активации.

С точки зрения автоматического применения модели нейронных сетей для прогнозирования изменения состояния объекта мониторинга необходимо выбрать приемлемую конфигурацию сети (числа входов, выходов, скрытых слоёв, количества нейронов в скрытых слоях, задание связей), вид функции активации, что является сложной вычислительной задачей по перебору возможных структур сети. Поэтому для применения нейронных сетей необходим опыт эксперта [61].

По той же причине для автоматического применения не подходят модели на базе цепей Маркова и классификационно-регрессионных деревьев.

3.4.3.2 Метод SSA

Рассмотрим сначала одномерный временной ряд $(f_i)_{i=1}^N$ [62]. Выберем n такое, что $0 < n < N-1$ – время жизни многомерной гусеницы. Пусть $\sigma = N-n+1$ – длина гусеницы. Построим последовательность n векторов в R^σ следующего вида:

$$Y^{(i)} \in R^\sigma, \\ Y^{(i)} = (f_{i+t-1})_{t=1}^\sigma$$

Обозначим $Z = (Y^{(1)}, \dots, Y^{(n)})$. Будем называть Z нецентрированной матрицей наблюдений, порожденной гусеницей со временем жизни n .

Рассмотрим ковариационную матрицу полученной выборки:

$$C = \frac{1}{n} ZZ^T.$$

Выполним ее svd-разложение:

$$C = V \Lambda V^T,$$

где $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$ – диагональная матрица собственных чисел, $V = (v^{(1)}, \dots, v^{(r)})$, $(v^{(i)})^T v^{(j)} = \delta_{ij}$ – ортогональная матрица собственных векторов.

Далее рассмотрим систему главных компонент:

$$U = V^T Z, U = (U^{(1)}, \dots, U^{(t)})^T.$$

После проведения анализа главных компонент обычно предполагается проведение операции восстановления исходной матрицы наблюдений по некоторому поднабору главных компонент, т.е. для $V' = (v^{(i_1)}, \dots, v^{(i_t)})$ и $U' = V'^T Z$ вычисляется матрица $Z' = V'U'$.

Далее восстанавливаются исходные данные последовательности. В одномерном случае i -ая компонента восстановленного ряда есть среднее значение по i -ой диагонали восстановленной матрицы наблюдений Z' .

Числовой ряд $(f_i)_{i=1}^{N+1}$ называется продолжением ряда $(f_i)_{i=1}^N$, если порождаемая им при гусеничной обработке выборка лежит в той же гиперплоскости, что и у исходного ряда $(f_i)_{i=1}^N$ [62]. Пусть у нас есть некоторый набор выбранных главных компонент i_1, i_2, \dots, i_t . Определим

$$w = \begin{pmatrix} v_s^{(i_1)} & v_s^{(i_2)} & \mathbf{L} & v_s^{(i_t)} \\ v_{2s}^{(i_1)} & v_{2s}^{(i_2)} & \mathbf{L} & v_{2s}^{(i_t)} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ v_t^{(i_1)} & v_t^{(i_2)} & \mathbf{L} & v_t^{(i_t)} \end{pmatrix}$$

и

$$V_* = \begin{pmatrix} v_1^{(i_1)} & v_1^{(i_2)} & \mathbf{L} & v_1^{(i_t)} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ v_{s-1}^{(i_1)} & v_{s-1}^{(i_2)} & \mathbf{L} & v_{s-1}^{(i_t)} \\ v_{s+1}^{(i_1)} & v_{s+1}^{(i_2)} & \mathbf{L} & v_{s+1}^{(i_t)} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ v_{2s-1}^{(i_1)} & v_{2s-1}^{(i_2)} & \mathbf{L} & v_{2s-1}^{(i_t)} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ v_{t-1}^{(i_1)} & v_{t-1}^{(i_2)} & \mathbf{L} & v_{t-1}^{(i_t)} \end{pmatrix}$$

Также положим $Q = (f_{N-s+2}^{(1)}, \dots, f_N^{(1)}, f_{N-s+2}^{(2)}, \dots, f_N^{(2)}, \dots, f_{N-s+2}^{(s)}, \dots, f_N^{(s)})^T$. Тогда прогнозируемые значения системы в точке $N+1$ вычисляются по формуле:

$$f_{N+1} = w(V_*^T V_*)^{-1} V_*^T Q.$$

3.4.3.3 Метод ARIMA

Логически модель ARIMA можно определить так [59]:

$$AR(p)+MA(q) \rightarrow ARMA(p,q) \rightarrow ARMA(p,q)(P,Q) \rightarrow ARIMA(p,q,r)(P,Q,R) \rightarrow \dots,$$

где $AR(p)$ – авторегрессионная модель порядка p ; $MA(q)$ – модель со скользящим средним порядка q .

Авторегрессионная модель имеет вид:

$$Y(t) = f_0 + f_1 Y(t-1) + f_2 Y(t-2) + \dots + f_p Y(t-p) + E(t),$$

где $Y(t)$ – зависимая переменная в момент времени t ; $f_0, f_1, f_2, \dots, f_p$ – оцениваемые параметры; $E(t)$ – ошибка от влияния переменных, которые не учитываются в данной модели.

Задача заключается в том, чтобы определить $f_0, f_1, f_2, \dots, f_p$. Их можно оценить различными способами. Правильнее всего искать их через систему уравнений Юла-Уолкера, для составления которой потребуется расчет значений автокорреляционной функции. Можно поступить более простым способом – посчитать их методом наименьших квадратов.

Модель со скользящим средним имеет вид [59]:

$$Y(t) = m + e(t) - w_1 e(t-1) - w_2 e(t-2) - \dots - w_q e(t-q),$$

где $Y(t)$ – зависимая переменная в момент времени t ; $w_0, w_1, w_2, \dots, w_q$ – оцениваемые параметры; $e(t)$ – случайное воздействие в момент времени t ; m – константа.

Этап идентификации модели ARIMA был описан в этом же пункте выше. Следующий, после идентификации, шаг состоит в оценивании параметров модели. Полученные оценки параметров используются на последнем этапе для того, чтобы вычислить новые значения ряда и построить доверительный интервал для прогноза. Процесс оценивания проводится по преобразованным данным (подвергнутым применению разностного оператора). До построения прогноза нужно выполнить обратную операцию (интегрировать данные). Таким образом, прогноз методологии будет сравниваться с соответствующими исходными данными.

3.4.3.4 Экспоненциальное сглаживание

Метод Хольта-Винтерса [63] основан на трехпараметрическом экспоненциальном сглаживании. При этом делается попытка учесть сезонные

составляющие, присутствующие в данных. Система уравнений, описывающих метод Хольта-Винтерса, выглядит следующим образом:

$$\begin{cases} \Omega_t = a \frac{Y_t}{S_{t-s}} + (1-a)(\Omega_{t-1} - T_{t-1}) \\ T_t = b(\Omega_t - \Omega_{t-1}) + (1-b)T_{t-1} \\ S_t = I \frac{Y_t}{\Omega_t} + (1-g)S_{t-s} \\ \hat{Y}_{t+p} = (\Omega_t + pT_t)S_{t-s+p} \end{cases},$$

где Ω – значения сглаженного ряда общего уровня; t – текущий момент времени; Y – значения временного ряда; S – значения сезонных колебаний; s – коэффициент сезонности; T – значения тренда (общее направление изменения входных данных); a, β, γ – постоянные сглаживания; p – число прогнозируемых значений; \hat{Y} – прогнозируемые значения временного ряда.

В приведенной системе уравнений первое из них описывает сглаженный ряд общего уровня. Дробь в этом уравнении служит для исключения сезонности из Y_t . После исключения сезонности алгоритм работает с данными, в которых нет сезонных колебаний. Появляются они в последнем уравнении, когда полученный прогноз умножается на сезонный коэффициент. Второе уравнение служит для оценки тренда. Третье уравнение определяет прогноз на p отсчетов по времени вперед.

3.4.4 Методы и алгоритмы нахождения характеристик выходных наборов данных системы административного мониторинга

3.4.4.1 Методы определения тренда

В работе [64] проводились исследования совокупности критериев, ориентированных на проверку гипотез отсутствия тренда в средних и дисперсиях. Критерии тренда и случайности предназначены для проверки гипотез о случайности расположения полученных выборочных данных, т.е. отсутствия взаимосвязи между значениями наблюдаемой случайной величины и их номерами в выборочной последовательности. Эти критерии

имеют широкую область применения, особенно при анализе временных рядов.

Ниже приведены результаты исследования.

Рассматриваемые критерии проверки гипотез об отсутствии тренда в средних, можно упорядочить по мощности (относительно линейного тренда) следующим образом:

Кокса-Стюарта > Бартелса > ранговой сериальной корреляции Вальда-Вольфовитца > автокорреляции > сериальной корреляции Вальда-Вольфовитца > Фостера-Стюарта.

Если рассматривать в качестве конкурирующих гипотез наличие периодического и смешанного тренда, то целесообразно рекомендовать использование критериев Бартелса, Вальда-Волфовитца (ранговый и неранговый вариант) и автокорреляции.

При исследовании мощности критериев обнаружения тренда в характеристиках рассеяния в качестве конкурирующих гипотез рассматривалась ситуация с линейным трендом в дисперсии. Исследовалась мощность критериев Фостера-Стюарта, Кокса-Стюарта, автокорреляции, ранговой сериальной корреляции Вальда-Волфовитца, сериальной корреляции Вальда-Вольфовитца и Бартелса. Оказалось, что критерии Фостера-Стюарта и Кокса-Стюарта, специально построенные для выявления тренда в дисперсии, весьма значительно превосходят по мощности остальные критерии. Для обнаружения тренда в дисперсии можно рекомендовать к применению критерии Фостера-Стюарта и Кокса-Стюарта, расположив их по предпочтению:

Фостера-Стюарта > Кокса-Стюарта.

По вышеописанным результатам можно сделать вывод, что для определения тренда в выходных наборах данных системы административного мониторинга нужно использовать два критерия: критерий Бартелса и Фостера-Стюарта.

Критерий Бартелса [65]. Пусть R_i – ранг i -го наблюдения в последовательности n наблюдений x_i . Бартелсом был предложен ранговый критерий случайности ряда, основанный на статистике:

$$B = \frac{\sum_{i=1}^{n-1} (R_i - R_{i+1})}{\sum_{i=1}^n (R_i - \bar{R})^2}.$$

Нормализованная статистика:

$$B^* = \frac{B - M(B)}{\sqrt{D(B)}} = \frac{B - 2}{\sqrt{\frac{5}{5n + 7}}}.$$

Если $|B^*| < u_{1-\alpha/2}$, то гипотеза случайности не отклоняется; $u_{1-\alpha/2}$ – квантиль стандартного нормального распределения.

Критерий Фостера-Стюарта [65]. Статистики критерия имеют вид:

$$S = \sum_{i=2}^n s_i; d = \sum_{i=2}^n d_i,$$

где $d_i = u_i - l_i$; $S_i = u_i + l_i$; $u_i = 1$, если $x_i > x_{i-1}, x_{i-2}, \dots, x_1$, иначе $u_i = 0$; $l_i = 1$, если $x_i < x_{i-1}, x_{i-2}, \dots, x_1$, иначе $l_i = 0$.

Статистика S используется для проверки наличия тренда в дисперсиях, а статистика d – для обнаружения тренда в средних. При отсутствии тренда величины $t = \frac{d}{f}$; $\tilde{t} = \frac{S - f^2}{l}$, приближенно подчиняются распределению Стьюдента с $\nu = n$ степенями свободы. Формулы для f и l применимы при $n > 50$, значения при $n \leq 50$ приведены в [66]:

$$l = \sqrt{2 \ln n - 3,4253}; f = \sqrt{2 \ln n - 0,8456}.$$

Нулевая гипотеза отсутствия тренда отклоняется с заданным уровнем значимости α , если $|t|, |\tilde{t}| > t_{1-\alpha/2}$.

3.4.4.2 Методы определения сезонной компоненты

В работе [67] используется автокорреляционная функция для исследования ряда на наличие периодической компоненты, сдвигая ряд на несколько отсчетов и сравнивая с самим собой. Автокорреляционная функция – это характеристика временного ряда, которая помогает находить

его повторяющиеся участки, скрытые из-за наложений шума или других помех. Для дискретного временного ряда X_1, X_2, \dots, X_n с известными матожиданием μ и дисперсией σ автокорреляционную функцию можно рассчитать по следующей формуле:

$$R(\omega) = \frac{1}{(n - \omega)\sigma^2} \sum_{t=1}^{n-\omega} [X_t - \mu][X_{t+\omega} - \mu],$$

где n – длина временного ряда, ω – текущая задержка во времени. Таким образом получим функция $R(\omega)$, зависящую от лагов (задержек во времени). Исследуя ее на экстремальные значения, получим значения периодов во временном ряде.

Анализ автокорреляционной функции позволяет выявить структуру ряда, т. е. определить присутствие в ряде той или иной компоненты. Так, если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тренд. Если наиболее высоким оказался коэффициент автокорреляции порядка m , то ряд содержит сезонную компоненту с периодичностью в m моментов времени. Если же ни один из коэффициентов автокорреляции не является значимым, то можно сделать одно из двух предположений:

1. Либо ряд не содержит тренда и сезонную компоненту, а его уровень определяется только случайной компонентой.
2. Либо ряд содержит сильный нелинейный тренд, для выявления которого нужно провести дополнительный анализ.

3.4.4.3 Методы определения случайной компоненты

Для определения значимости случайной компоненты будет использоваться совокупность характеристик моделей хранения – уровень ОУ, количество связей с ОУ по показателям, количество терминальных ОУ.

В результате анализа различных структур хранения данных при решении конкретных задач административного мониторинга:

1. Если объект учета имеет самый низкий уровень, то он имеет значимую случайную компоненту.

2. Если ОУ имеет меньше 10 терминальных ОУ, то он имеет значимую случайную компоненту.

3. Если количество связей с ОУ по показателям меньше 3, то он имеет значимую случайную компоненту.

Для определения количества терминальных ОУ и количества связей с ОУ по показателям будет использоваться смешанный обход по дереву, которым представлена логическая структура модели хранения данных. Уровень ОУ может быть определен нисходящим обходом по дереву, которым представлена логическая структура модели хранения.

3.4.4.4 Методы определения выбросов

Во временных рядах наибольшее распространение получил метод Ирвина, основанный на определении λ — статистики [68]. При его использовании выявление аномальных наблюдений производится по формуле:

$$I_i = \frac{|y_i - y_{i-1}|}{s_y},$$

где y_i — текущее значение во временном ряду, которое проверяется на аномальность;

y_{i-1} — значение во временном ряду идущее перед тем, которое проверяется на аномальность;

σ_y — среднеквадратическое отклонение уровней ряда от их среднего уровня.

Если расчетное значение превысит уровень критического $\lambda_{кр}$ (с заданным уровнем точности и числом наблюдений) (таблица 3.1), то расчетное значение признается аномальным.

Таблица 3.1 – Табулированные значения λ_i

Число наблюдений	$\lambda_{кр}$	
	0,95	0,99
2	2,8	3,7
3	2,2	2,9
10	1,5	2,0
20	1,3	1,8
30	1,3	1,7
50	1,1	1,6
100	1,0	1,5

3.4.5 Нахождение оптимального набора характеристик временного ряда для автоматического выбора алгоритма прогнозирования

В результате проведенного исследования был сформирован набор потенциальных признаков временного ряда, которые могут повлиять на выбор прогнозной модели: 11 характеристик временных рядов и 3 характеристики модели хранения данных системы административного мониторинга (период учета будет использоваться как вспомогательная характеристика для нахождения других характеристик). Эти признаки позволят определить максимальный размер обучающей выборки необходимой для формирования факторного плана.

3.4.5.1 Построение обучающей выборки

Для построения обучающей выборки необходимо сначала определить значения, которые будет принимать каждый признак в отдельности. В терминологии планирования экспериментов входные переменные и структурные допущения, составляющие модель, называются факторами, а выходные показатели откликами. В данном эксперименте признаки

временного ряда являются факторами [69]. Также различают управляемые и неуправляемые факторы. К управляемым факторам из сформированного выше набора потенциальных признаков относятся: длина ряда, наличие или отсутствие тренда, наличие или отсутствие сезонной компоненты, наличие или отсутствие случайной компоненты, количество пропущенных значений, количество выбросов, количество вспомогательных временных рядов, количество терминальных объектов учета, связанных по общему показателю. Такие признаки как количество вспомогательных временных рядов, уровень объекта учета и количество связей с терминальными объектами учета по общему показателю относятся к неуправляемым факторам, так как их влияние на отклик получается только из-за особенностей предметной области и конкретной задачи, поэтому в рамках текущего эксперимента использовать значения этих факторов будет невозможно. Оставшиеся 4 признака относятся также к неуправляемым факторам, но их значения можно будет получить после генерации временного ряда, поэтому они будут учитываться в эксперименте и соответственно при составлении обучающей выборки. Для того чтобы обучающая выборка состояла из временных рядов, содержащих любое сочетание управляемых признаков, необходимо построить полный факторный план. Для этого определим для каждого управляемого признака (фактора) два значения – максимальное и минимальное, которые он может принять:

1. Длина ряда – 1000 и 50.
2. Наличие или отсутствие тренда – 1 и 0.
3. Наличие или отсутствие сезонной компоненты – 1 и 0.
4. Степень влияния случайной компоненты – 1 и 0.
5. Количество пропущенных значений – 5% от длины ряда и 0.
6. Количество выбросов – 5% от длины ряда и 0.
7. Количество терминальных объектов учета, связанных по общему показателю – 100 и 0.

Таким образом, получаем, что необходимо сгенерировать 2^7 временных рядов, чтобы построить полный факторный план. Если во временном ряде присутствует тренд, то случайным образом выбирается его образующая функция: логарифмическая, степенная (невысокого порядка), функция обратной пропорциональности. Если во временном ряде присутствует сезонная компонента, то прибавляется тригонометрическая функция синуса. Сезонная компонента также может и умножаться на временной ряд, но в рамках данного эксперимента будем пользоваться аддитивной сезонной компонентой. При необходимости можно будет внести мультипликативную сезонную компоненту, как еще один классификационный признак. Если во временном ряде присутствует случайная компонента, то для каждого значения временного ряда выбирается по нормальному закону распределения целое случайное число из диапазона 0..99, затем оно масштабируется (нормализуется) в зависимости от текущего максимального и минимального значений во временном ряде и после этого прибавляется к уже имеющемуся значению. Так же как и с сезонной компонентой при необходимости можно внести, случайную компоненту с равномерным законом распределения в качестве дополнительного классификационного признака. Это связано с тем, что на практике нормальный закон распределения встречается намного чаще, чем равномерный. Если во временном ряде есть пропущенные значения, то они помечаются специальным символом. Если во временном ряде есть выбросы, то для каждого выброса случайным образом определяются: будет ли он складываться или вычитаться из текущего значения. Если во временном ряде есть терминальные объекты учета связанные по общему показателю, то эти значения временных рядов, представленных показателями терминальных объектов учета, будут складываться между собой и делиться на их общее количество.

После того как будут сформированы 2^7 временных рядов, представляющие собой полный факторный план, необходимо для каждого временного ряда установить какой из методов прогнозирования, будет

работать лучше на каждом временном ряде. Для этого делаем прогноз на 10% значений от длины временного ряда каждым методом прогнозирования. Затем сравниваем результат с реальными (сгенерированными) значениями, используя в качестве критерия среднеквадратичную ошибку. Алгоритм с наименьшей среднеквадратичной ошибкой является лучшим для конкретного временного ряда.

В итоге получаем обучающую выборку, представляющую собой две матрицы: первая матрица T размерами 128×11 (7 управляемых и 4 неуправляемых признаков), в каждой строке которой находятся значения признаков, характеризующих временной ряд, вторая матрица A вектор-столбец размерами 1×128 , состоящая из значений, закрепленными за методами прогнозирования, которые работают наилучшим образом на временном ряде. Теперь необходимо осуществить поиск, оптимального набора классификационных признаков.

3.4.5.2 Алгоритм полного перебора для поиска оптимального набора классификационных признаков

Для нахождения оптимального набора классификационных признаков был разработан специальный алгоритм. Данный алгоритм перебирает каждое сочетание классификационных признаков, решает задачу нахождения весов для классификационных признаков при помощи метода наименьших квадратов, описанного выше, и запоминает процент успешно классифицированных рядов для найденных весов. По завершению алгоритм находит сочетание (сочетания) классификационных признаков с самым высоким процентом.

Ниже приведено более подробное описание алгоритма полного перебора для поиска оптимального набора классификационных признаков (Рисунок 3.2). На входе имеются две матрицы T и A , описанные в предыдущем пункте.

1. Для каждого числа i от 0 до $127 (2^7)$ выполнять пункты 2-14.

2. Перевести число i в двоичное представление.
3. Скопировать матрицу T в M . Записать в k 0 (счетчик количества управляемых признаков в текущем сочетании признаков).
4. Для первых семи разрядов двоичного числа i повторить пункт 5.
5. Если на месте текущего разряда двоичного числа находится 0, то удалить из матрицы M соответствующий номеру разряда столбец (берем из обучающей выборки только те признаки, которые соответствуют данному сочетанию i), иначе увеличить k на 1.
 6. Транспонировать матрицу M результат записать в M^T .
 7. Умножить матрицы $M^T \times M$.
 8. Найти обратную матрицу от произведения $(M^T \times M)^{-1}$ методом Гаусса.
 9. Умножить обратную матрицу от произведения $(M^T \times M)^{-1}$ на матрицу M^T , а получившуюся матрицу затем умножить на матрицу A , результат записать в матрицу вектор-столбец W (Матрица W имеет размер $1 \times (k+4)$, в ней содержатся значения весов для текущего сочетания признаков).
 10. Для первых семи разрядов двоичного числа i повторить пункт 11.
 11. Если на месте текущего разряда двоичного числа находится 0, то добавить соответствующую нулевую строку в матрицу W (добавляем нулевые веса для признаков, которые не учитываются в текущем сочетании).
 12. Умножить матрицу T на матрицу W , результат записать в матрицу вектор-столбец A' (Размер матрицы A' 1×128).
 13. Привести значения вектор-столбца матрицы A' к ближайшим значениям, закрепленным за методами прогнозирования (Определяем лучшие методы прогнозирования для T с текущим сочетанием признаков).
 14. Сравнить значение в каждой строке матрицы A' с соответствующим значением строки матрицы A , высчитать процент совпавших значений, итог записать в i ячейку результирующего массива R .
 15. Найти номер (номера) сочетания признаков в массиве R с наибольшим значением.

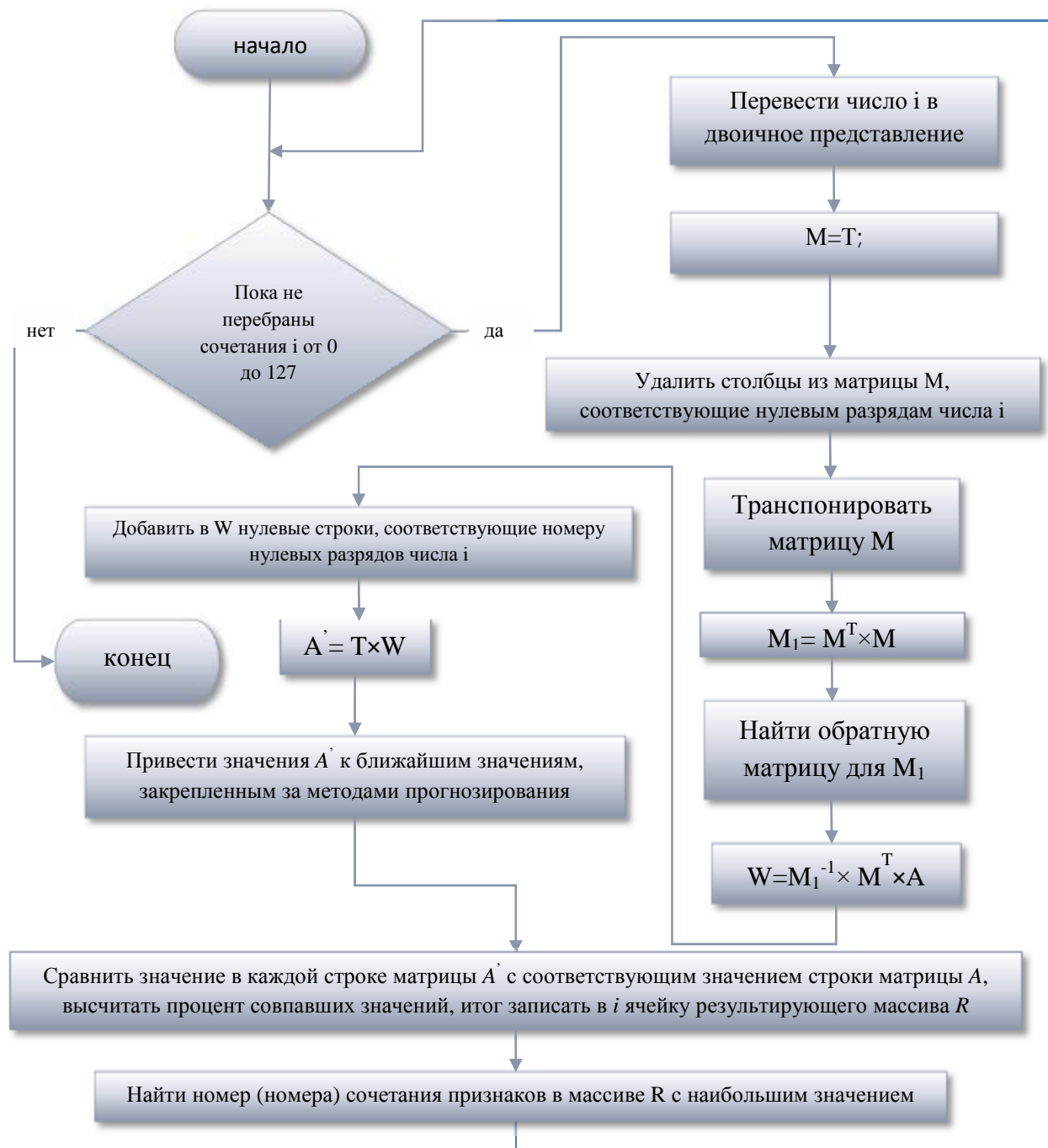


Рисунок 3.2 – Алгоритм поиска оптимального набора признаков

3.4.5.3 Получение классификационной функции для автоматического выбора алгоритма прогнозирования

По приведенному выше алгоритму был произведен вычислительный эксперимент для трех методов прогнозирования: SSA, ARIMA, тройное экспоненциальное сглаживание. В результате эксперимента был достигнут максимум классификационной функции оптимальности на сочетании

признаков: наличие или отсутствие тренда, наличие или отсутствие сезонной компоненты, наличие или отсутствие случайной компоненты, количество выбросов, количество терминальных объектов учета, связанных по общему показателю. Для данного сочетания признаков количество правильно классифицированных рядов равно 128 из 128, поэтому можно со 100 % вероятностью автоматически выбрать наилучший метод прогнозирования среди трех вышеописанных методов.

Значения, закрепленные за методами прогнозирования SSA, ARIMA, тройное экспоненциальное сглаживание были -5, 0, 5 соответственно. Классификационная функция с учетом полученных весов для найденного сочетания признаков выглядит следующим образом:

$$F(X) = -1.25x_1 - 13.75x_2 + 3.75x_3 + 1.25x_4 - 0.0275x_5,$$

где x_1, x_2, x_3, x_4, x_5 – классификационные признаки: наличие или отсутствие тренда, наличие или отсутствие сезонной компоненты, наличие или отсутствие случайно компоненты, количество выбросов, количество терминальных объектов учета, связанных по общему показателю – соответственно.

По приведенным весам и набору классификационных признаков можно сделать следующие выводы:

1. Метод SSA лучше других методов прогнозирует ряд, если в нем есть сезонная компонента.
2. Менее чувствительные к выбросам являются методы прогнозирования как ARIMA и тройное экспоненциальное сглаживание.
3. Лучше всего прогнозирует ряд, где отсутствует детерминированная составляющая, метод тройного экспоненциального сглаживания.

3.5 Выводы по третьей главе

1. На основании модели структуры и содержания выходных отчетных форм, представленной предложенным в работе языком описания шаблона

выходных отчетных форм, был разработан алгоритм формирования шаблона отчета с возможностью поуровневой настройкой объектов учета.

2. Были описаны формальные методы классификации, кластеризации и прогнозирования при решении задач анализа данных административного мониторинга и разработаны алгоритмы контроля данных.

3. На основании модели правила анализа для задачи прогнозирования и анализа методов решения задачи прогнозирования выявлена необходимость решения задачи автоматического выбора наилучшего метода прогнозирования состояния объекта административного мониторинга из заданного множества. Для ее решения был разработан специальный алгоритм. В результате его работы была получена классификационная функция с помощью, которой можно осуществлять автоматический выбор наилучшего метода прогнозирования состояния объекта мониторинга из заданного семейства методов прогнозирования для конкретного состава и структуры показателей в текущий момент решения задачи административного мониторинга.

4 Реализация и исследование прототипа средств адаптации автоматизированной системы административного мониторинга

4.1 Структурная схема подсистем генерации отчетов, анализа и контроля данных

На основании модели процесса переработки данных и методики автоматизации процессов анализа и контроля данных была разработана структурная схема подсистем генерации отчетов, анализа и контроля данных в системе административного мониторинга (рисунок 4.1) [5].

В подсистему настройки анализа и контроля данных поступает вектор управления системой мониторинга, который представляет собой: типы ОУ, показатели, нормативные значения, период выборки данных, источник данных и т.д. В подсистеме настройки анализа и контроля данных происходит инициализация определяемых пользователем переменных, являющихся аргументами функций, полученных при моделировании анализа и контроля. Результатами подсистемы настройки анализа и контроля данных являются правила анализа и критерии контроля данных.

В подсистему настройки отчетов поступает вектор управления системой мониторинга, который представляет собой структуру отчета: отчетные ОУ, показатели, стили визуализации и т.д. В подсистему настройки отчетов Результатами подсистемы настройки отчетов являются шаблоны отчетов.

Подсистема анализа данных использует данные, полученные из хранилища или подсистемы генерации отчетов. Его результаты в подавляющем большинстве случаев могут рассматриваться как самостоятельные информационно-аналитические ресурсы, пригодные для принятия управленческих решений. Однако, для некоторых задач анализа, например, задачи прогнозирования, где результатом являются значения

показателей, возможно использование этих результатов как объекта контроля.

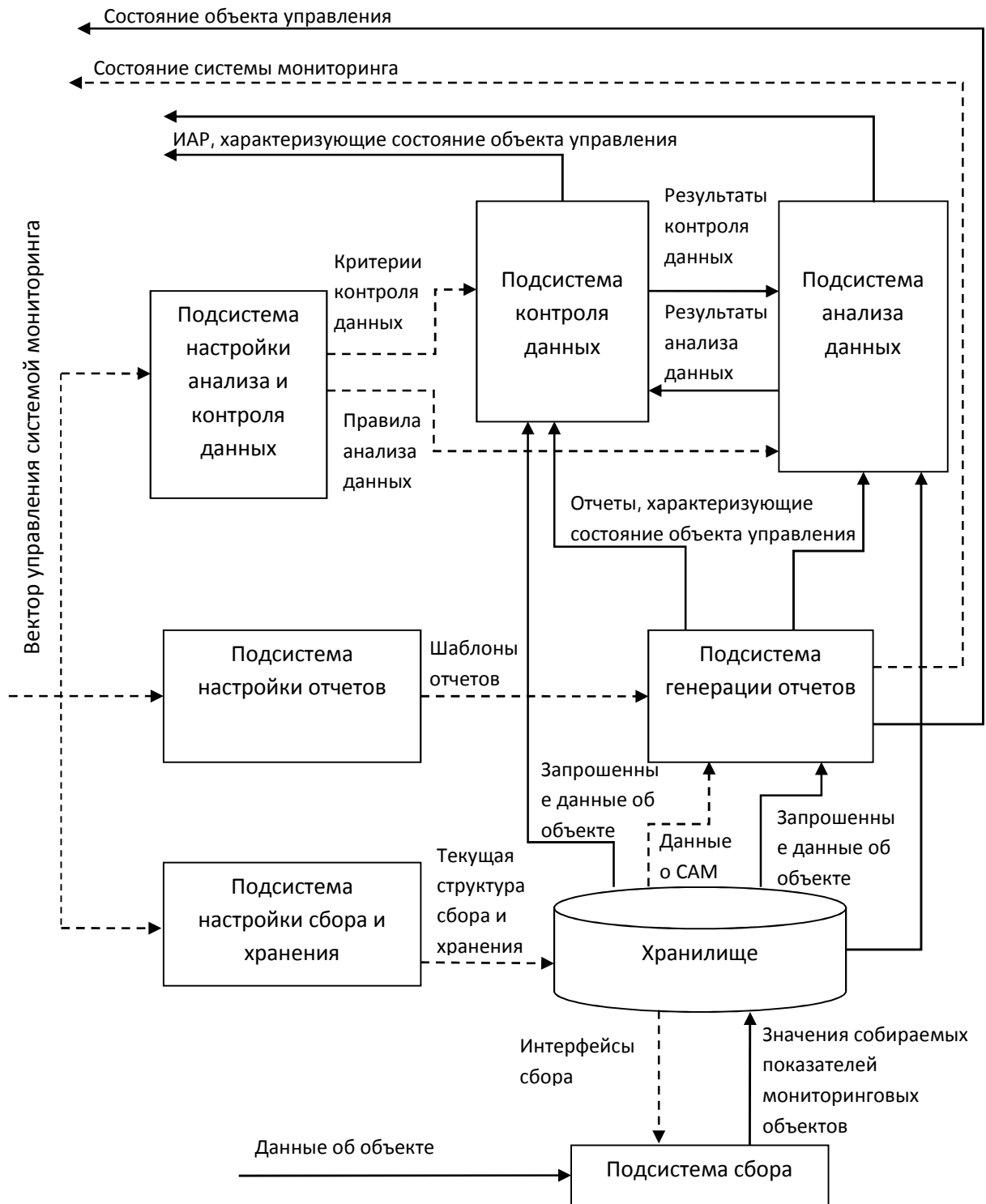


Рисунок 4.1 – Структурная схема подсистем генерации отчетов, анализа и контроля данных

Исходными данными для подсистемы контроля данных могут являться результаты подсистемы генерации отчетов или данные из хранилища [5]. Результаты контроля данных могут использоваться как информационно-аналитические ресурсы, пригодные для принятия управленческих решений. Также по результатам контроля данных можно проводить анализ данных, например, прогнозирование для объектов учета, у которых показатели близки к выходу от нормативных значений.

В подсистему генерации отчетов поступает выборка данных из хранилища для конкретного настроенного шаблона отчетов. После этого происходит формирование выходного отчета. Выходные отчеты характеризуют состояние объекта управления, а также состояния системы мониторинга.

4.2 Диаграммы состояний диалогового интерфейса пользователя

Интерфейс подсистемы анализа и контроля данных должен позволять настраивать правила анализа и критерии контроля данных, производить настройку задачи анализа и контроля данных, выполнять запуск настроенных задач и просматривать результаты их решения. Соответственно, в составе подсистемы анализа и контроля должны быть реализованы три диалоговых интерфейса.

- 1 Интерфейс настройки правил анализа и критериев контроля данных.
- 2 Интерфейс настройки задач анализа и контроля данных. На этом же интерфейсе будет выполняться и запуск настроенных задач.
- 3 Интерфейс вывода результатов работы подсистемы.

Подсистема формирования выходных отчетов должна иметь следующие интерфейсы:

- интерфейс настройки шаблона выходной отчетной формы;
- интерфейс донастройки типовых шаблонов отчетов;
- интерфейс настройки стиля визуализации (модели структуры и правил формирования) отчета;

4.2.1 Интерфейс настройки подсистемы генерации отчетов

Задание структуры шаблона отчетов ведется по уровням, где пользователь последовательно задает настройки текущего уровня: типы объектов учета, объекты учета, их показатели, необходимость вести аналитический подсчет (обработку данных) в случае, если показатели числовые и прочее. Поэтому интерфейс настройки структуры шаблона отчетов рационально сделать интерфейсом типа «wizard». В таком случае задача разбивается на ряд интуитивно понятных шагов:

Ожидание выбора множества объектов учета для визуализации на текущем уровне. Для удобства пользователя и ускорения отбора на интерфейсе необходимо обеспечить возможность обобщающей выборки: показывать все объекты учета данного уровня, показывать объекты учета определенного типа (из типов, присутствующих на уровне), показать определенные экземпляры объектов учета. Подобный интерфейс позволяет выбрать, например, объекты учета разного типа, если эти типы существуют на рассматриваемом уровне.

Ожидание выбора показателей для выбранного множества объектов учета, по которым необходимо показывать значения на данном уровне. На этом этапе определяется: будет ли отчет простым (без «аналитики» (обработки данных)) или сложным. Ведение анализа значений показателей и вычисление агрегирующих функций возможны только на числовых показателях. Поэтому возможны следующие варианты включения значений показателей на данном уровне: использовать числовые показатели (с возможностью агрегации как на данном уровне, так и во всем дереве), использование любых показателей (только вывод значения, без возможности агрегации и другой аналитики), не использовать значения показателей в отчете на данном уровне (для отчетов, требующих только выявления родственных связей объектов учета).

Ожидание выбора метода анализа данных для числовых показателей. На данном этапе при доступности только данных объектов учета одного

поколения, группировка для проведения агрегации возможна по следующим наборам объектов учета: группировка данных по всем объектам учета на данном уровне; группировка данных по типам (агрегирующие и другие анализирующие функции выдают значения для каждой группы объектов учета данного уровня, разделенных по типу), группировка данных по одному родителю.

Ожидание выбора объектов учета, для которых необходимо выводить значения показателей.

Ожидание выбора объектов учета, потомки которых будут обработаны на следующем уровне иерархии.

Интерфейс донастройки типовых шаблонов представляет собой набор уникальных генерируемых интерфейсов. Генерация интерфейса должна осуществляться на основе анализа типового шаблона отчета и выявления множеств исходных данных, необходимых для построения и задаваемых пользователем. Соответственно, каждый генерируемый интерфейс должен будет обеспечивать ввод информации (различной для каждого типового шаблона), необходимой для донастройки отчета:

- информация о типах объектов учета;
- информация об экземплярах объектов учета;
- информация о показателя объектов учета.

Интерфейс настройки стиля визуализации, также как и интерфейс создания шаблона, также рационально реализовать в виде wizard-образного интерфейса. Данный вид интерфейса представляет собой последовательность сменяющих друг друга форм и обработчиков. Каждая форма позволяет пользователю ввести данные об одном из атрибутов стиля и передает их обработчику. Создание стиля можно в любой момент отменить, а также вернуться последовательно на несколько шагов назад. Интерфейс будет включать четыре основных формы, после прохождения которых заполненный атрибут отправляется обработчиком в базу данных.

Перечень данных форм определен на основании анализа модели структуры и правил формирования выходных отчетов (на рисунке 4.2 основные формы обозначены прямоугольниками):

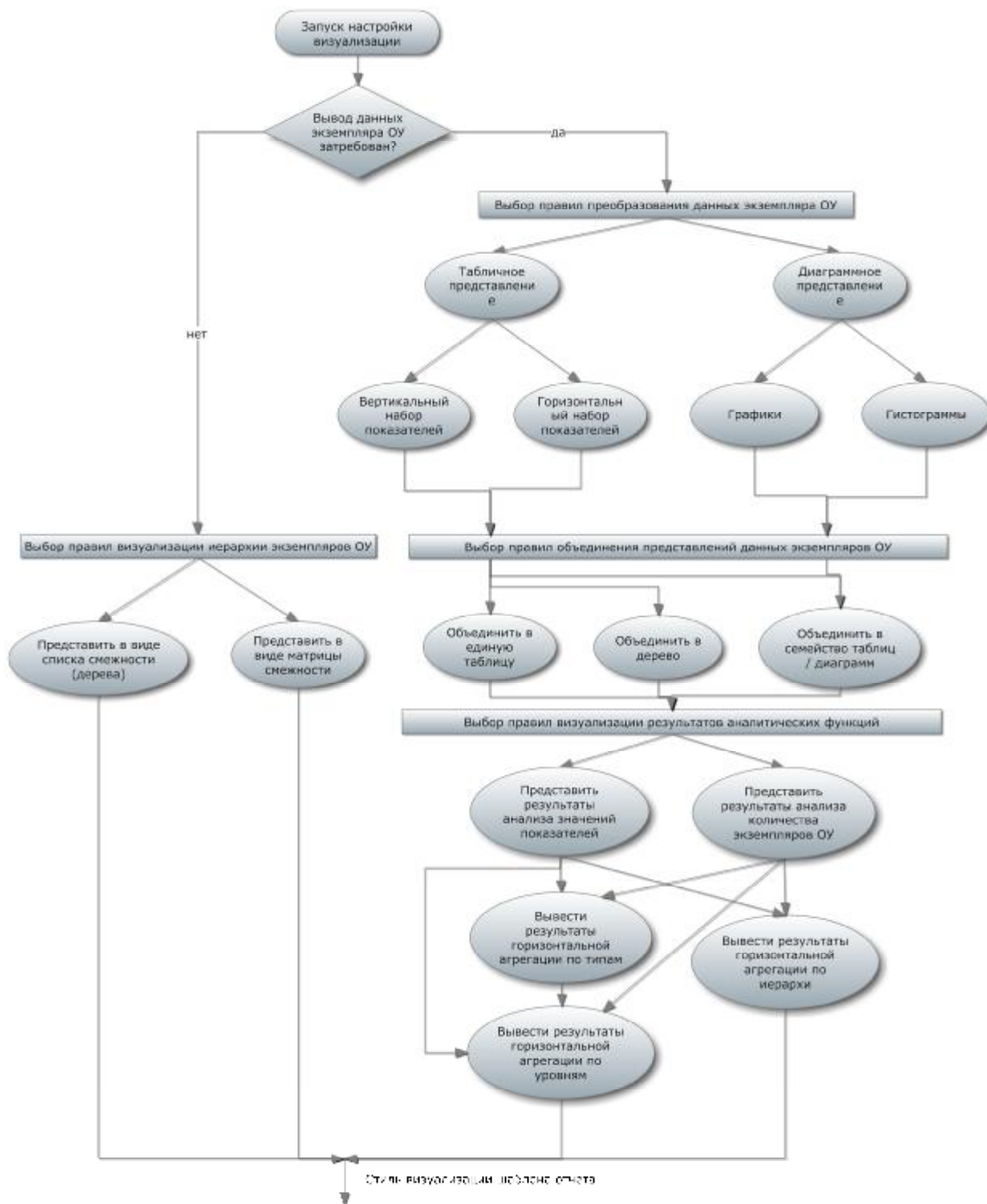


Рисунок 4.2 – Интерфейс настройки стиля визуализации

- форма выбора правил визуализации иерархии экземпляров объектов учета;
- форма выбора правил преобразования данных объекта учета;
- форма выбора правил объединения представлений данных экземпляров объектов учета;
- форма выбора правил визуализации результатов аналитических функций (функций обработки данных).

На рисунке 4.2 также представлена детализация состояний перечисленных выше основных форм интерфейса.

Интерфейс настройки периода выборки данных для отчета представляет собой интерфейс типа SDI, в котором пользователю предлагается выбрать начальную и конечную дату из календаря.

Интерфейс вывода отчетов определяется стилем визуализации для заданной структуры шаблона отчетов. Также для обеспечения возможности экспорта полученного отчета на данном интерфейсе должен присутствовать элемент управления для выбора выходного формата файла для экспорта выходной отчетной формы.

4.2.2 Интерфейс настройки правил анализа и критериев контроля данных

На этом интерфейсе происходит добавление новых или редактирование существующих правил анализа и критериев контроля данных. В отличие от добавления, редактирование позволяет изменять лишь некоторые настройки правил анализа и критериев контроля данных. Также на этом интерфейсе доступно удаление правил анализа и критериев контроля данных.

При добавлении происходит выбор типа правила или критерия контроля данных, в зависимости от которого вызывается соответствующий интерфейс настройки:

- интерфейс настройки правил анализа данных для задачи классификации;

- интерфейс настройки правил анализа данных для задачи кластеризации;
- интерфейс настройки правил анализа данных для задачи прогнозирования;
- интерфейс настройки критериев контроля максимального и минимального значений;
- интерфейс настройки критериев контроля разницы между наибольшим и наименьшим значениями;
- интерфейс настройки критериев контроля среднего арифметического;
- интерфейс настройки критериев контроля выбросов.

Рассмотрим действия пользователя на каждом из интерфейсов.

Интерфейс настройки правила анализа данных для задачи классификации.

- 1 Ожидание выбора типа объекта учета.
- 2 Ожидание выбора показателей данного типа, на основании которых будут определяться классы объектов учета.
- 3 Ожидание выбора обучающего множества объектов учета.
- 4 Ожидание ввода значений зависимого показателя (класса) для выбранного типа объектов учета.
- 5 Ожидание сохранения настроек правила анализа данных.

При редактировании правил анализа данных для задачи классификации, будет вызван этот же интерфейс, на котором будут доступны только 3, 4 и 5 пункты.

Интерфейс настройки правила анализа данных для задачи кластеризации.

- 1 Ожидание выбора типа объекта учета.
- 2 Ожидание выбора показателей данного типа, на основании которых будет идти разбиение на кластеры.
- 3 Ожидание выбора метрики расстояний.
- 4 Ожидание ввода количества кластеров.

5 Ожидание сохранения настроек правила анализа данных.

При редактировании правил анализа данных для задачи кластеризации, будет вызван этот же интерфейс, на котором будут доступны только 3, 4 и 5 пункты.

Интерфейс настройки правил анализа данных для задачи прогнозирования.

1 Ожидание выбора объекта учета.

2 Ожидание выбора периодического показателя у данного объекта учета.

3 Ожидание ввода количества значений периодического показателя, на основании которых будет рассчитываться прогноз.

4 Ожидание ввода количества периодов, на которое необходимо сделать прогноз.

5 Ожидание сохранения настроек правила анализа данных.

При редактировании правил анализа данных для задачи прогнозирования, будет вызван этот же интерфейс, на котором будут доступны только 3, 4 и 5 пункты.

Интерфейс настройки критериев контроля максимального и минимального значений:

1 Ожидание выбора типа объекта учета.

2 Ожидание выбора показателя данного типа объекта учета.

3 Ожидание ввода максимального и минимального допустимых значений для данного показателя.

4 Ожидание выбора типа критерия контроля: четкий или нечеткий.

5 Ожидание ввода количества частей, на которое разбивается контролируемый диапазон значений (только для нечетких критериев).

6 Ожидание ввода для каждой части ее процентного значения от общего контролируемого диапазона и символического имени, характеризующего объекты учета, у которых значения показателей попали в эту часть (только для нечетких критериев).

7 Ожидание сохранения результатов.

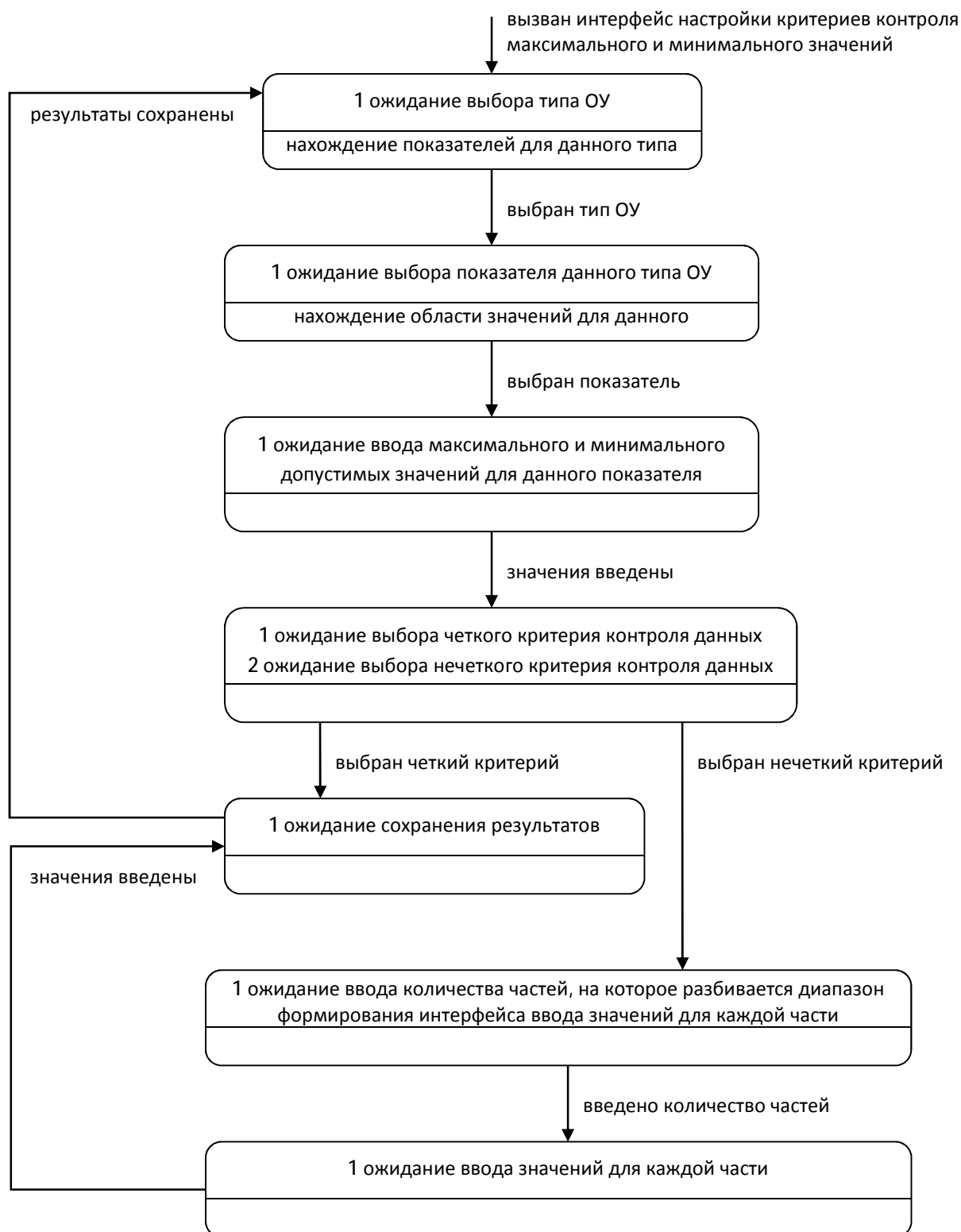


Рисунок 4.3 – Интерфейс настройки критериев контроля максимального и минимального значений

При редактировании критериев контроля максимального и минимального значений будет вызван этот же интерфейс, на котором будут доступны только пункты с 3 по 7.

Интерфейсы настройки других критериев контроля данных будут отличаться только 3 пунктом. Логика диалога интерфейса настройки критериев контроля максимального и минимального значений (в качестве примера) изображена ниже на рисунке 4.3.

4.2.3 Интерфейс настройки задачи анализа и контроля данных

На данном интерфейсе происходит настройка новых и редактирование, удаление, запуск существующих задач анализа и контроля данных. При запуске задач анализа и контроля данных может потребоваться выборка данных из базы данных системы мониторинга.

При добавлении вызывается интерфейс настройки задач анализа и контроля, характеризующийся следующим набором состояний.

1 Ожидание выбора правила или критерия анализа и контроля данных.

2 Ожидание ввода названия задачи анализа и контроля данных

3 Ожидание выбора режима запуска задачи анализа и контроля данных.

4 Ожидание выбора источника данных (выборка данных из базы данных, выбор настроенного отчета, выбор задачи анализа данных (только для задачи контроля данных)).

5 Ожидание ввода значения периода выборки данных (только для запросов выборки из базы данных и подсистемы генерации отчетов).

6 Ожидание выбора способа представления результатов.

7 Ожидание сохранения результатов настройки.

При редактировании задач анализа и контроля данных будет вызван тот же интерфейс, на котором будут доступны пункты со 2 по 7.

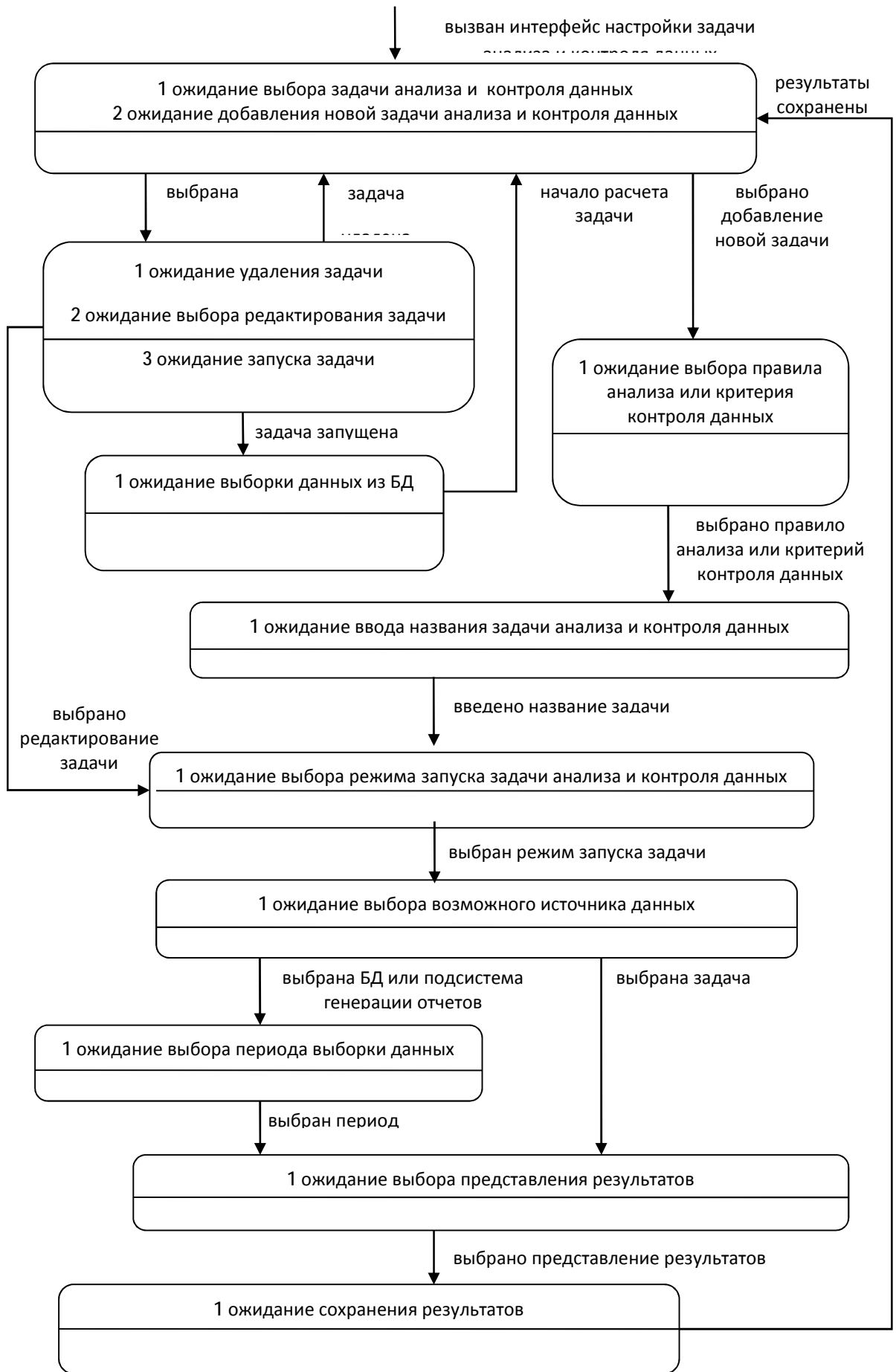


Рисунок 4.4 Интерфейс настройки задачи анализа и контроля данных

Логика диалога интерфейса настройки задачи анализа и контроля данных приведена на рисунке 4.4.

4.3 Архитектура программных средств подсистем генерации отчетов, анализа и контроля данных

В качестве основы архитектуры реализации была выбрана классическая трехуровневая архитектура [70] [71], включающая компоненты хранилища, логики и представления. Однако в свете современных представлений о проектировании, будем использовать видоизменную архитектуру, более подходящую под концепцию MVC («model-view-controller»). От трехуровневой она отличается тем, что в ней выделен дополнительный связующий компонент – «контроллер». Контроллер является звеном между «представлением» и «логикой» (model в англоязычной литературе) системы, посредством которого и существует возможность произвести разделение между ними. Контроллер получает данные от пользователя и передаёт их в компонент «логика». Кроме того, он получает сообщения от логики, и передаёт их в «представление». Контроллер следит за переданными данными от пользователя и проводит их первичный анализ, в результате которого определяет: передать данные модулю логики или обработать ошибочный ввод без привлечения модуля логики, отправив «представлению» сообщение о некорректном вводе.

Кроме того, контроллер обязан определять тип данных, полученных от модуля логики (есть ли это готовый результат, отсутствие результата, либо сообщение об ошибке) и передавать информацию в модуль представления.

Фактически связка контроллера и представления и определяют интерфейс пользователя. Контроллер стоит на входе обработки пользовательской активности, а представление – на выходе.

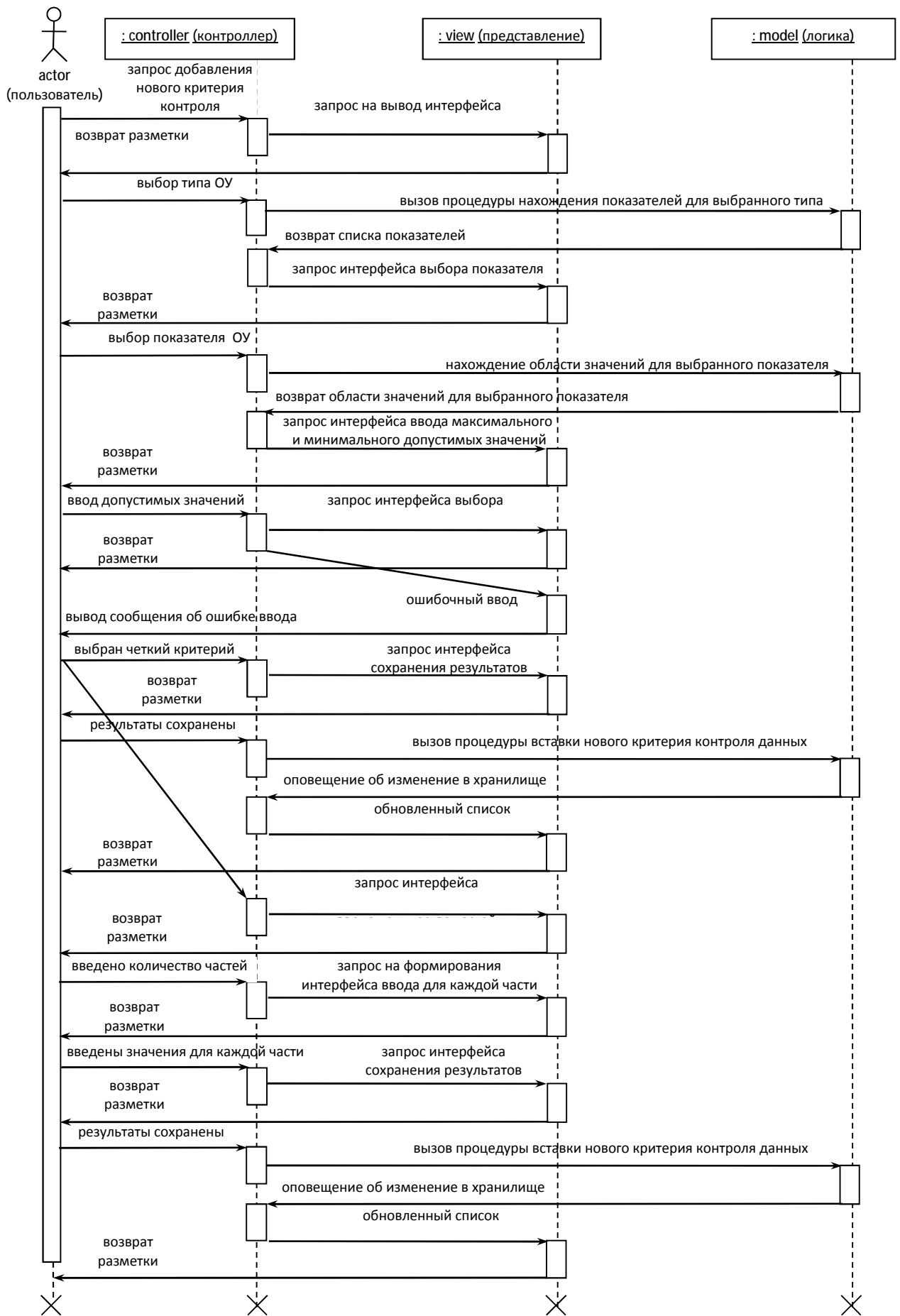


Рисунок 4.5 – Диаграмма последовательности

На модуль логики данные поступают в формате, удобном для построения запросов к хранилищу, реализованному согласно модели структуры хранения. Модуль логики на основе полученных данных формирует необходимые запросы к хранилищу и обрабатывает результат.

Данная архитектура позволяет строить алгоритмы, генерирующие все перечисленные выше виды интерфейса. Поскольку схема отработки ввода новых данных в указанной архитектуре будет принципиально отличаться для интерфейсов обработки данных, то представим на рисунке 4.5 цикл отработки генерации интерфейса добавления критерия контроля максимального и минимального допустимых значений.

Сначала пользовательский запрос на добавление нового критерия контроля данных поступает в контроллер. Контроллер передает в представление запрос на вывод интерфейса выбора типа объекта учета. Представление возвращает пользователю разметку интерфейса. После чего пользователь может выбирать тип объекта учета. Сообщение о выборе типа объекта учета поступает в контроллер. Контроллер передает запрос в логику на вызов процедуры нахождения показателей для выбранного типа. Логика возвращает в контроллер список показателей. Затем контроллер передает в представление запрос на вывод интерфейса выбора показателей.

В первом случае пользовательский запрос выполнен без привлечения логики, а во втором логика производила поиск списка показателей.

4.4 Исследование прототипа и оценка эффективности функционирования средств адаптации подсистем генерации отчетов, анализа и контроля данных автоматизированной системы административного мониторинга

В первой главе было сказано, что вследствие уникальности модели хранения данных на физическом уровне, реализованной в адаптивной системе административного мониторинга, ни один из существующих

генераторов отчетов и ни одна из технологий генерации не могут быть напрямую перенесены в рассматриваемую систему, поэтому оценку показателей эффективности разработанной подсистемы генерации отчетов будем производить с результатами оценки трудоемкости реализации неавтоматизированной методики построения отчетов. Рассматривать импорт данных из АСАМ для построения отчетов в какие-либо другие системы исключим из рассмотрения, поскольку время создания шаблона импорта для конкретной задачи мониторинга будет сопоставимо со временем разработки подсистемы генерации отчетов АСАМ. При изменении задачи мониторинга потребуется почти полностью переписать шаблон импорта данных.

Сопутствующими эффектами от разработки прототипа средств адаптации системы административного мониторинга будут являться социальный эффект, выражающийся в улучшении условий труда и снижении механической нагрузки на работника, уменьшении вредного влияния компьютера и организационный эффект, заключающийся в изменении процессов переработки информации.

Для наглядности представим этапы, в соответствии с которыми проводилась оценка, хронологически (Рисунок 4.).

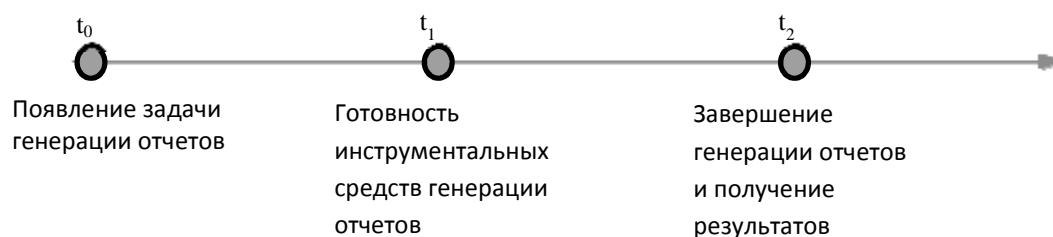


Рисунок 4.6 – Этапы процедуры генерации отчетов

Этап подготовки для АСАМ представляет собой создание шаблона отчета и связанного с ним стиля визуализации для новой задачи генерации отчета. В дальнейшем при появлении задачи генерации отчетов будет использоваться уже имеющийся шаблон. В неавтоматизированной методике генерации отчетов подготовка подразумевает доступ к данным и анализ задачи.

Таблица 4.1 – Этапы процедуры генерации отчетов

	Неавтоматизированная методика	АСАМ
t_0-t_1	Анализ задачи Доступ к данным ($T_{подг_ручн}$)	Задание нового стиля визуализации Задание новой структуры шаблона Сохранение структуры шаблона в хранилище ($T_{подг_асм}$)
t_1-t_2	Построение отчетов Обработка данных ($T_{обр_ручн}$)	Выборка данных из БД САМ Обработка данных ($T_{обр_асм}$)

Сравнение эффективности подсистемы анализа и контроля данных адаптируемой системы мониторинга будем проводить с эффектом от импорта данных в системы класса business intelligent, позволяющим настроиться на любую задачу анализа и контроля данных адаптируемой САМ, из-за наличия большой библиотеки различных методов анализа и контроля данных, и с результатами оценки трудоемкости реализации неавтоматизированной методики анализа и контроля данных.

Для наглядности представим этапы, в соответствии с которыми проводилась оценка, хронологически (рисунок 4.7).

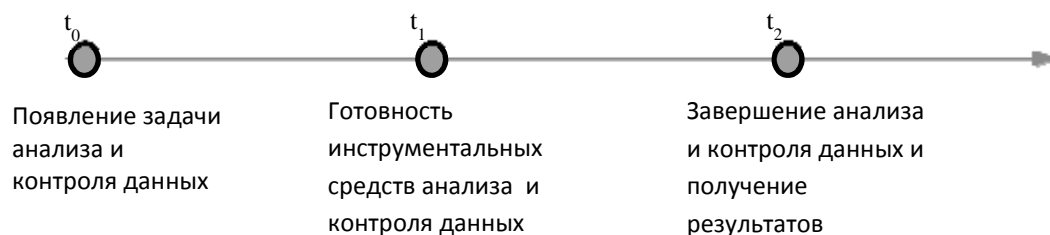


Рисунок 4.7 – Этапы процедуры анализа и контроля данных

Таблица 4.2 – Этапы процедуры анализа и контроля данных

	Неавтоматизированная методика	Система класса business intelligent	АСАМ
t_0-t_1	Анализ задачи Доступ к данным ($T_{наст_ручн}$)	Создание шаблона импорта данных из САМ Импорт данных из САМ Настройка параметров для анализа и контроля данных ($T_{наст_импр}$)	Настройка правил анализа и критериев контроля данных Настройка общей задачи анализа и контроля данных Выборка данных из подсистемы генерации отчетов или БД САМ Выборка данных из подсистемы генерации отчетов или БД САМ ($T_{наст_асм}$)
t_1-t_2	Анализ и контроль данных ($T_{анко_ручн}$)	Анализ и контроль данных ($T_{анко_импр}$)	Анализ и контроль данных ($T_{анко_асм}$)

При оперативном прогнозировании показателей объектов учета возникает необходимость быстрого нетривиального выбора метода прогнозирования и нужных параметров модели для самого метода, поэтому задача прогнозирования существенно отличается от других задач анализа и контроля данных, и оценку эффективности ее реализации необходимо произвести отдельно как для систем класса business intelligent, так и для неавтоматизированной методики прогнозирования.

Для наглядности представим этапы, в соответствии с которыми проводилась оценка, хронологически (рисунок 4.8).

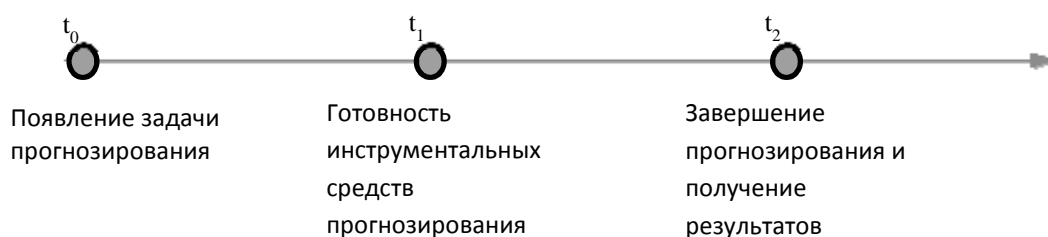


Рисунок 4.8 – Этапы процедуры прогнозирования

Таблица 4.3 – Этапы процедуры прогнозирования

	Неавтоматизированная методика	Система класса business intelligent	АСАМ
t_0-t_1	Доступ к данным Анализ временного ряда ($T_{\text{напр_ручн}}$)	Создание шаблона импорта данных из САМ Импорт данных из САМ Анализ временного ряда Выбор метода прогнозирования Настройка параметров для прогнозирования ($T_{\text{напр_импр}}$)	Настройка правила анализа для задачи прогнозирования Настройка общей задачи прогнозирования Выборка данных из подсистемы генерации отчетов или БД САМ Анализ временного ряда Выбор метода прогнозирования ($T_{\text{напр_асм}}$)
t_1-t_2	Построение прогноза ($T_{\text{прог_ручн}}$)	Построение прогноза ($T_{\text{прог_импр}}$)	Построение прогноза ($T_{\text{прог_асм}}$)

4.4.1 Оценка временных показателей по указанным интервалам

Для оценки временных и трудовых затрат средств адаптации подсистем генерации отчетов, анализа и контроля данных системы административного мониторинга будем использовать данные, полученные при разработке опытного образца и его внедрении.

Оценку временных и трудовых показателей разработки шаблона импорта данных из АСАМ для системы класса business intelligent получим с помощью одной из алгоритмических моделей оценки стоимости разработки программного обеспечения The Constructive Cost Model 2 (COCOMO 2). Данный способ оценки использует регрессионную формулу с параметрами, собранными из базы данных на основе статистической информации о многих проектах аналогичного класса.

Поскольку уровень опыта разработчиков проектов разный, то примем за расчетный средний класс проекта (Intermediate/Semi-detachedmode) со средними по размеру командами со смешанным опытом разработки и

смешанными требованиями. Выбранный класс определит расчетные параметры.

Согласно данной модели трудозатраты рассчитываются следующим образом:

$$Z_{\text{спец_разр}} = a_b (\text{KLOC})^b_b \text{ (человеко-месяцев);}$$

$$T_{\text{спец_разр}} = c_b (Z_{\text{спец_разр}})^d_b * 24 * 30 \text{ (человеко-часов).}$$

Коэффициенты a_b , b_b , c_b и d_b перечислены ниже:

$$a_b=3.0; b_b=1.12; c_b=2.5; d_b=0.35.$$

Общий объем разработки шаблона импорта данных из АСАМ класса business intelligent составил 0.02 тысяч функциональных строк.

$$Z_{\text{шабл_разр}} = 3.0 * (0.02)^{1.12} = 12.32 \text{ (человеко-месяцев)}$$

$$T_{\text{шабл_разр}} = 2.5 * (12.32)^{0.35} * 24 * 30 = 34 \text{ (человеко-часов)}$$

Итоговые расчетные оценки для задачи генерации отчетов:

$$T_{\text{подг_ручн}} = 90 \text{ человеко-часов;}$$

$$T_{\text{подг_асм}} = 12 \text{ человеко-часов;}$$

$$T_{\text{обр_ручн}} = 80 \text{ человеко-часов;}$$

$$T_{\text{обр_асм}} = 6 \text{ человеко-час.}$$

Итоговые расчетные оценки для задач анализа и контроля данных (этап настройки для системы класса business intelligent складываются из общего объема разработки шаблона импорта $T_{\text{шабл_разр}}$ и настроек параметров для задачи анализа и контроля $T_{\text{наст_анко}}$):

$$T_{\text{наст_ручн}} = 60 \text{ человеко-часов;}$$

$$T_{\text{наст_импр}} = T_{\text{шабл_разр}} + T_{\text{наст_анко}} = 34 + 8 = 42 \text{ человеко-часа;}$$

$$T_{\text{наст_асм}} = 8 \text{ человеко-часов;}$$

$$T_{\text{анко_ручн}} = 60 \text{ человеко-часов};$$

$$T_{\text{анко_импр}} = 6 \text{ человеко-часа};$$

$$T_{\text{анко_асм}} = 6 \text{ человеко-часов.}$$

Итоговые расчетные следующие для задачи прогнозирования (этап настройки для системы класса business intelligent складываются из разработки шаблона импорта для задачи прогнозирования $T_{\text{шабл_прог}}$ и анализа временного ряда, выбора метода прогнозирования, настройки параметров для прогнозирования $T_{\text{наст_прог}}$):

$$T_{\text{напр_ручн}} = 30 \text{ человеко-часов};$$

$$T_{\text{напр_импр}} = T_{\text{шабл_прог}} + T_{\text{наст_прог}} = 10 + 14 = 24 \text{ человеко-часа};$$

$$T_{\text{напр_асм}} = 3 \text{ человеко-часов};$$

$$T_{\text{прог_ручн}} = 20 \text{ человеко-часов};$$

$$T_{\text{прог_импр}} = 2 \text{ человеко-часа};$$

$$T_{\text{прог_асм}} = 2 \text{ человеко-часов.}$$

Представим полученные суммарные данные на диаграмме (Рисунок 4.-4.11).

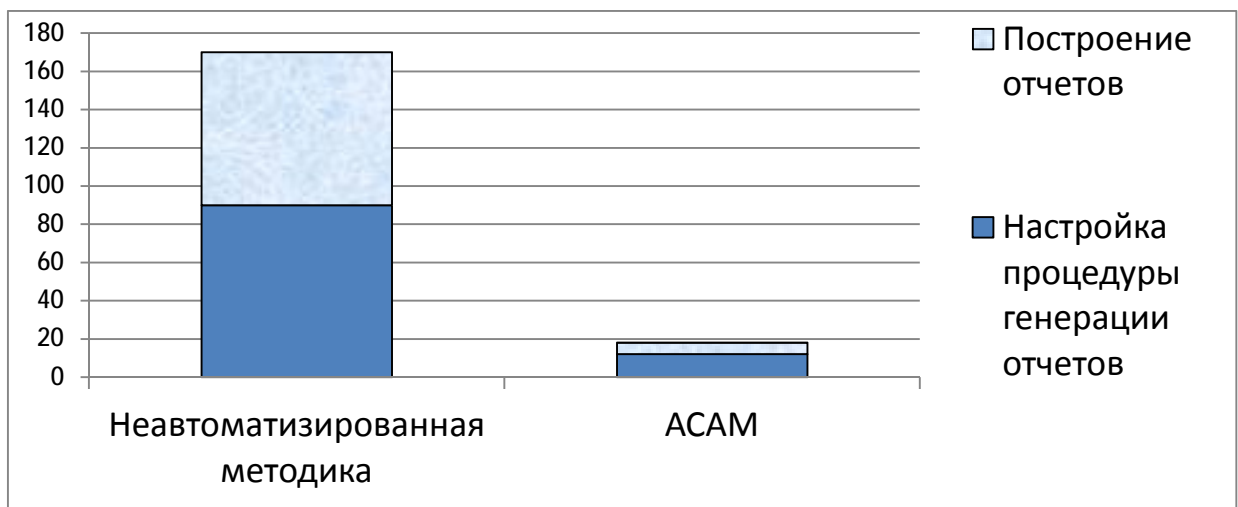


Рисунок 4.9 – Затраты времени на одну итерацию генерации отчетов после изменения требований к процедуре мониторинга, чел*час

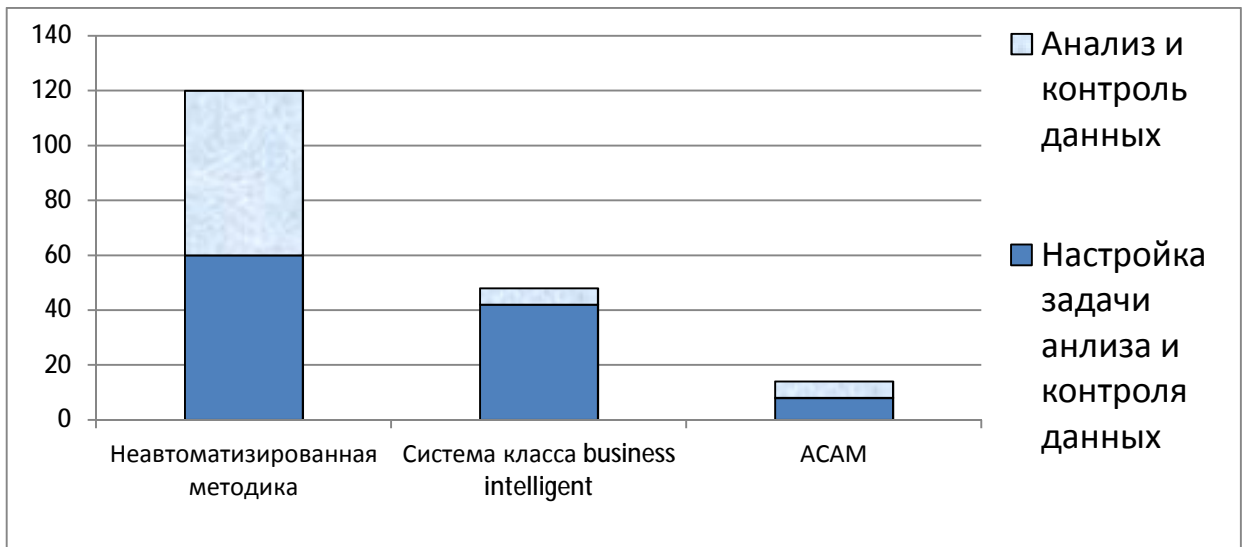


Рисунок 4.10 — Затраты времени на одну итерацию анализа и контроля данных после изменения требований к процедуре мониторинга, чел*час

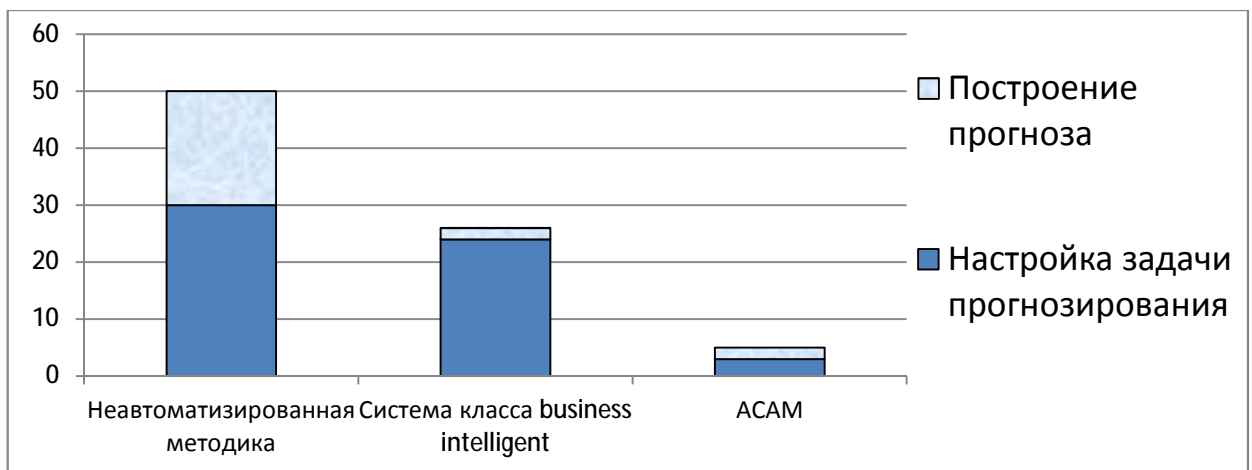


Рисунок 4.11 — Затраты времени на одну итерацию прогнозирования после изменения требований к процедуре мониторинга, чел*час

Затраты времени на внедрение и эксплуатацию подсистемы анализа и контроля данных АСАМ в среднем соответствует внедрению и эксплуатации системы класса business intelligent, с разработанным шаблоном импорта данных для конкретной задачи. Обычно эти затраты связаны с закупкой нового оборудования и обучением персонала. Количество и мощность оборудования, а также количество и квалификация людей, требуемых для работы с системой мониторинга при грубом расчете можно принять за

равные у конкурирующих рассматриваемых вариантов, поскольку система предоставляет «дружественный» интерфейс, соответствующий уровню уверенного пользователя ЭВМ. В связи с данным комментарием расчеты по оценке стоимостных затрат на внедрение и эксплуатацию «вынесем за скобки» и исключим из сравнения.

4.4.2 Примеры

Результаты работы подсистемы анализа и контроля данных могут быть представлены в табличном виде, в текстовом виде, в виде различного рода графиков.

Результаты решения задач классификации, как и результаты решения задач кластеризации удобнее всего представлять в табличном виде, где в столбцах обозначаются классы или кластеры, по которым осуществлено разбиение множества объектов учета, а в строках – объекты учета. Однако для наглядности, если в задаче кластеризации количество измерений (показателей), по которым идет разбиение, равно двум или трем, то можно построить графики. На этих графиках показатели будут откладываться на координатных осях, а объекты учета, принадлежащие одному кластеру, будут изображаться одной фигурой или цветом.

Результаты решения задачи прогнозирования лучше всего представляются в графическом виде, где на оси абсцисс откладываются значения периодов учета, а на оси ординат значение показателя для данного периода. Потом эти значения соединяются между собой последовательно или строится некая аппроксимирующая функция. Также результаты решения задачи прогнозирования можно представить в табличном виде.

Результаты контроля данных нагляднее всего представлять в табличном виде, где в столбцах находятся названия множества, характеризующие успешность прохождения контроля данных, а в строках сами объекты учета. Однако в некоторых случаях, особенно с нечеткими критериями контроля данных, наиболее эффективным будет вариант

представления с помощью текстовых конструкций (лингвистических переменных).

Возможным способы представления результатов решения задач анализа и контроля данных приведены в таблице 4.4.

Таблица 4.4 – Представление результатов анализа и контроля данных

Название задачи анализа и контроля данных	Табличный способ	Графический способ	Текстовый способ
Классификация	+	-	-
Кластеризация	+	+	-
Прогнозирование	+	+	-
Четкий контроль данных	+	-	+/-
Нечеткий контроль данных	+	-	+

Приведем пример решения задачи получения сводных данных на примере построения простого отчета: «Посчитать сумму по мероприятиям 4.1, 5.4, 7.7, 7.12, 7.14, 8.5, выделенную на госконтракты, заключенные в 2011 году».

В терминах системы мониторинга: «Посчитать количество объектов учета типа «госконтракт» за период учета 2011 год, родителями которых являются экземпляры ОУ типа «мероприятия» 4.1, 5.4, 7.7, 7.12, 7.14, 8.5. Для каждого такого ОУ типа «госконтракт» передать своему родителю значение показателя «финансирование» (Рисунок 4.12).

Проведем разбиение ОУ типа «мероприятия» на два кластера по результатам вышеописанного отчета. Кластеризация ОУ будет вестись по показателям «финансирование» и «заключено госконтрактов», в качестве метрики расстояний между ОУ используется евклидово расстояние. На рисунке 4.13 представлена матрица разбиения по кластерам.

Объекты учета типа "мероприятие"	Профинансиро вано, тыс. руб.	Заклучено госконтрактов
	2011	2011
4.1 Разработка механизмов реализации градостроительной политики (включая, финансово-правовые механизмы) с учетом перехода на стратегическое планирование и программно-целевое бюджетирование	2322.6	1
5.4 Разработка нормативно-технических документов в области проектирования, строительства и безопасной эксплуатации подземных сооружений	23053	9
7.7 Научно-технические и методические разработки в области применения новых материалов и технологий	37134.086	6
7.12 Модернизация системы сертификации и нормативно-технической документации, включая создание системы энергосберегающих стандартов в строительном комплексе	43823.505	14
7.14 Разработка и реализация системы безопасной эксплуатации уникальных и технически сложных зданий и сооружений, в том числе большепролетных, с использованием современных автоматизированных технических средств	38394.804	7
8.4 Научно-методические разработки в области обеспечения безопасности зданий и сооружений и энергосберегающего домостроения	45568.469	17

Рисунок 4.12 – Пример построения простого отчета

На результатах того же отчета можно провести контроль минимального допустимого значения для ОУ типа «мероприятие» по показателю «заклучено госконтрактов» за период учета 2011 год (Рисунок 4.14). Минимально допустимое значение равно 3.

Объекты учета типа "мероприятие"	Кластер 1	Кластер 2
4.1	Не принадлежит	Принадлежит
5.4	Не принадлежит	Принадлежит
7.7	Принадлежит	Не принадлежит
7.12	Принадлежит	Не принадлежит
7.14	Принадлежит	Не принадлежит
8.4	Принадлежит	Не принадлежит

Рисунок 4.13 – Разбиение ОУ типа «мероприятия» на два кластера

Объекты учета типа "мероприятие"	Контроль показателя "заключено контрактов"
	2011
4.1	Не пройден
5.4	Пройден
7.7	Пройден
7.12	Пройден
7.14	Пройден
8.4	Пройден

Рисунок 4.14 – Контроль минимального допустимого значения для ОУ типа «мероприятие» по показателю «заключено госконтрактов»

4.5 Выводы по четвертой главе

1. В качестве основы архитектуры программных средств подсистем генерации отчетов, анализа и контроля данных была выбрана модификация трехуровневой архитектуры, включающая компоненты хранилища, логики и представления и контроллер.

2. Реализован прототип подсистем генерации отчетов, анализа и контроля данных. В ходе внедрения результатов работы получены оценки эффективности средств адаптации прототипа, выражающейся в сокращении времени на настройку задач переработки данных: в 7 раз для задачи получения сводных данных по сравнению с неавтоматизированной методикой, в 5 раз для задачи анализа и контроля данных по сравнению с системами класса business intelligent, без учета задачи прогнозирования; и в 8 раз для задачи прогнозирования по сравнению с системами класса business intelligent.

ЗАКЛЮЧЕНИЕ

В процессе диссертационного исследования были получены следующие результаты:

1. Был проведен анализ особенностей административного мониторинга средств и методов его организации. По результатам анализа выявлено, что системам поддержки принятия решений и/или лицу принимающему решения, необходимо оперативно получать информацию из различных источников с последующей ее переработкой до необходимого уровня абстракции о состоянии объекта управления, в условиях изменяющихся требований к структуре и составу показателей. Вследствие чего должна обеспечиваться возможность оперативной адаптации автоматизированной системы административного мониторинга к этим требованиям.

2. Решение задачи оперативной адаптации автоматизированной системы административного мониторинга может быть достигнуто за счет формализованного представления задач получения, анализа и контроля первичных и сводных данных в терминах специальной модели хранения данных административного мониторинга.

3. В ходе анализа задач получения сводных данных, анализа и контроля данных была разработана модель процесса переработки данных, обеспечивающая возможность повышения абстракции представления объекта мониторинга и унификацию на этой основе представления исходных данных и результатов процесса переработки.

4. Для решения задач получения, анализа и контроля первичных и сводных данных системы административного мониторинга были разработаны формализованные модели правил анализа и критериев контроля данных, структуры и содержания выходных отчетных форм в терминах модели хранения данных системы административного мониторинга.

5. На основании модели процесса переработки данных была разработана методика построения процессов автоматизации анализа и контроля данных с возможностью выбора источника данных.

6. На основании модели структуры и содержания выходных отчетных форм, представленной предложенным в работе языком описания шаблона выходных отчетных форм, был разработан алгоритм формирования шаблона отчета с возможностью поуровневой настройкой объектов учета.

7. На основании модели правила анализа для задачи прогнозирования и анализа методов решения задачи прогнозирования выявлена необходимость решения задачи автоматического выбора наилучшего метода прогнозирования состояния объекта административного мониторинга из заданного множества. Для ее решения был разработан специальный алгоритм. В результате его работы была получена классификационная функция с помощью, которой можно осуществлять автоматический выбор наилучшего метода прогнозирования состояния объекта мониторинга из заданного семейства методов прогнозирования для конкретного состава и структуры показателей в текущий момент решения задачи административного мониторинга.

8. Реализован прототип подсистем генерации отчетов, анализа и контроля данных. В ходе внедрения результатов работы получены оценки эффективности средств адаптации прототипа, выражающейся в сокращении времени на настройку задач переработки данных: в 7 раз для задачи получения сводных данных по сравнению с неавтоматизированной методикой, в 5 раз для задачи анализа и контроля данных по сравнению с системами класса business intelligent, без учета задачи прогнозирования; и в 8 раз для задачи прогнозирования по сравнению с системами класса business intelligent.

9. Результаты диссертационного исследования использованы при выполнении пяти НИР и внедрены в учебный процесс кафедры «Информационные системы» ФГБОУ ВПО «Госуниверситет – УНПК».

Список литературы

1. Единая система стандартов автоматизированных систем управления [Текст]: ГОСТ 24.103-84 Взамен гост 16084-75. - 1984.
2. Павлов, А.А. Основы системного анализа и проектирования АСУ [Текст] / А.А. Павлов, С.Н. Гриша, В.Н. Томашевский. - Киев: Выща школа, 1991. - с.367
3. Коськин, А.В. Информационно-аналитические ресурсы для управления организационно-техническими системами: монография [Текст] / А.В. Коськин. – М: Машиностроение-1, 2006. – 208 с.
4. Кравцова, Н.А. Методика организации процессов сбора, хранения и обработки данных в автоматизированной адаптивной системе административного мониторинга [Текст] - Новосибирск: НГТУ, 2010. - сс.139-141
5. Лукьянов, П.В. Моделирование процессов анализа и контроля данных административного мониторинга при автоматизации управления организационно-техническими системами [Текст] / П.В. Лукьянов, А.И. Фролов // Информационные системы и технологии. – 2012. – №6(74). – С. 84-90.
6. Ковалевский, С.С. Создание систем мониторинга реализации Федеральных целевых программ [Текст] / С.С. Ковалевский, В.В. Кульба. - М.: Синтег, 2006. - с.160
7. Фролов, А.И. Административный мониторинг как элемент процесса организации управления в организационно-технических системах [Текст]: ИПУ РАН / А.И. Фролов, И.С. Константинов. // Управление развитием крупномасштабных систем (MLSD'2011): Материалы Пятой международной конференции, 2011 - Т. 2. - сс.383-386
8. Черняк, Ю.И. Системный анализ в управлении экономикой [Текст] / Ю.И. Черняк. - М.: Экономика, 1975. - с.191

9. Сладкова, О.Б. Информационный мониторинг: теоретико-методологические основы [Текст] / О.Б. Сладкова. - Москва: МГУКИ, 2002. - с.65
10. Сенченко, П.В. Мониторинг социально-экономического развития территорий в контексте информационного обеспечения системы управления по результатам [Текст]: АлтГТУ/ П.В. Сенченко, А.А. Сидоров. //Ползуновский альманах, 2009 - сс.293-299
11. Рычихина, Э.Н. Показатели эффективности управленческого мониторинга муниципального образования [Текст] : ВШЭ/ Э.Н. Рычихина. //Вопросы государственного и муниципального управления, 2010
12. Рычихина, Э.Н. Мониторинг как общая функция управления [Текст] / Э.Н. Рычихина. - Ухта: УГТУ, 2007. - с.47
13. Боровкова, Т.И. Мониторинг развития системы образования Часть 1. Теоретические аспекты [Текст] / Т.И. Боровкова, И.А. Морев. - Владивосток: Дальневосточного университета, 2004. - с.150
14. Информационно-аналитическая система мониторинга выполнения приоритетных национальных проектов [Электронный ресурс] //Группа компаний ЛАНИТ. - 2010. - Режим доступа: http://www.lanit-consulting.ru/products/oracle/indust_sol/
15. Информационно-аналитическая Система мониторинга деятельности сети подведомственных бюджетных учреждений (ИАС МДБУ) в социально значимых отраслях Республики Татарстан [Электронный ресурс] //ООО "БАРС Груп". - Режим доступа: <http://mincult.tatar.ru/rus/docs/monitoring-system.htm>
16. Информационно-аналитическая система мониторинга показателей результатов деятельности организаций промышленного и оборонно-промышленного комплексов Министерства промышленности и торговли Российской Федерации [Электронный ресурс] //ФГУП "ЦНИИ "Центр". - Режим доступа: http://www.cniicentr.ru/scientifically_centre/information_system/

17. Информационно-аналитическая система "Экологический мониторинг" [Электронный ресурс] // Web-сервер администрации Калужской области. - 2020. - Режим доступа: <http://ecoanalyt.ru/ecoweb/>
18. Парк, Д. Сбор данных в системах контроля и управления [Текст] / Д. Парк, С. Маккей. - М.: Группа ИДТ, 2006. - с.504
19. Авдулов, П.В. Введение в теорию принятия решений [Текст] / П.В. Авдулов. - М.: ИУНХ, 1977. - сс.8-9
20. Александрова, А.П. Организация административного мониторинга социальных программ на региональном и местном уровнях [Текст] / ред. А.П. Александрова. - М: Фонд "Институт экономики города", 2002. - с.52
21. Перегудова, Ф.И. Системное проектирование АСУ хозяйства области [Текст] / ред. Ф.И. Перегудова. - М.: Статистика, 1977. - с.10
22. Маслаков, М.А. Процессно-ориентированные информационные системы [Текст] / М.А. Маслаков, В.Н. Якимов. // Автоматизация и современные технологии, 2009 - сс.17-22
23. Nilsson, J. Применение DDD и шаблонов проектирования [Текст] / J. Nilsson. - Вильямс.
24. Кравцова Н.А. Автоматизация процессов сбора и хранения данных при проведении административного мониторинга: Дис. ... канд. тех. наук. Орел. 2012. - 164 с.
25. Константинов, И.С. Модель хранения данных в адаптивной автоматизированной системе административного мониторинга [Текст] / И.С. Константинов, А.И. Фролов, Н.А. Кравцова // Информационные системы и технологии. – 2010. – №4(60). – С. 66-73.
26. Кравцова, Н.А. Место и функции адаптивной системы мониторинга в автоматизированной системе управления [Текст] / Н.А. Кравцова, А.И. Фролов // Информационные технологии в науке, образовании и производстве. ИТНОП-2010: материалы IV-й Международной научно-технической конференции, 2010. – Т. 2. – С. 71-75.

- 27.ГОСТ Р 51170-98. Качество служебной информации. Термины и определение [Текст]. Введ. 1998–05–12. – М.: Изд-во стандартов, 1998. – 7с.
- 28.Большая советская энциклопедия [Электронный ресурс]. – Научное издательство «Большая Российская энциклопедия», 2001. – Режим доступа: <http://slovari.yandex.ru/документ/БСЭ/Документ>.
- 29.Jackson M.A. Principles of program design. – London: Academic Press, 1975. – 297 p.
- 30.Дюк, В.А. Применение технологий интеллектуального анализа данных в естественнонаучных, технических и гуманитарных областях [Текст]/ В.А. Дюк, А.В. Флегонтов, И.К. Фомина // Известия Российского Государственного Педагогического Университета им. А.И. Герцена. Естественные и точные науки. Вып. 138, 2011. С. 77-84.
- 31.Барсегян, А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP [Текст]/ А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – 2-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2007.– 384 с.: ил. CD-ROM, ISBN 5-94157-991-8.
- 32.Константинов, И.С. Формализация задач анализа и контроля данных в автоматизированной адаптивной системе административного мониторинга / И.С. Константинов, А.И. Фролов, П.В. Лукьянов // Информационные системы и технологии: материалы Международной научно-технической Интернет-конференции, апрель-май 2011 г., г. Орел – В 3-х т. Т. 1 / под общ. ред. д-ра техн. наук проф. И.С. Константинова. – Орел: ФГОУ ВПО «Госуниверситет-УНПК», 2011. – С. 31-37.
- 33.Автоматизированные системы. Основные положения [Текст]: Методические указания - М.: Изд-во стандартов, 1998. - с.7
- 34.Архипенков, С.Я. Хранилища данных [Текст] / С.Я. Архипенков, Д.В. Голубев, О.Б. Максименко. - М.: Диалог-МИФИ, 2002. - с.528

35. Коннолли, Т. Базы данных: проектирование, реализация и сопровождение. Теория и практика: изд. 2-е; пер. с англ. [Текст] / Т. Коннолли, К. Бегг, А. Страчан. - М.: "Вильямс", 2003. - с.1440
36. Бурков, В.Н. Введение в теорию управления организационными системами [Текст] / В.Н. Бурков, Н.А. Коргин, Д.А. Новиков. - М.: Либриком, 2009. - с.264
37. Codd, E.F. Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate [Текст]: Technical Report/ E.F. Codd, S.B. Codd, C.T. Salley. - San Jose: Codd & Date Inc, 1993. - с.31
38. Елманова, Н. Введение в OLAP-технологии Microsoft [Текст] / Н. Елманова, А. Федоров. - М.: Диалог-МИФИ, 2002. - с.272
39. Parsaye, K. OLAP and Data Mining: Bridging the Gap [Текст]: Database Programming and Design/ K. Parsaye., 1998 - №10. - сс.30-37
40. Han, J. OLAP Mining: An Integration of OLAP with Data Mining [Текст] / J. Han. - IFIP, 1997. - с.18
41. Фролов А.И., Кравцова Н.А. Вопросы алгоритмизации процессов генерации адаптируемых диалоговых интерфейсов в системе административного мониторинга // Компьютерные науки и технологии: Компьютерные науки и технологии: сборник трудов Второй Международной научно-технической конференции. 3-5 октября 2011, г. Белгород. – Белгород: ООО «ГиК», 2011. – С 131-135.
42. Лукьянов П.В. Моделирование структуры выходных отчетных форм в системе административного мониторинга [Текст] // «Информационно-телекоммуникационные системы и технологии» (ИТСиТ-2012) : Материалы Всероссийской молодежной конференции, г. Кемерово: Кузбас. гос. техн. ун-т им. Т.Ф. Горбачева, 20-22 сентября 2012 г, сс. 40-41.
43. ISO/IEC 14977:1996 «Information technology – Syntactic metalanguage – Extended BNF». – 2001. – 12 p.

44. Лукьянов, П.В. Модели правил анализа и критериев контроля данных в системе административного мониторинга [Текст] / И.С. Константинов, А.И. Фролов, П.В. Лукьянов // «Информационные системы и технологии», №4(72), июль-август 2012, сс. 70-77.
45. Лукьянов, П.В. Алгоритм формирования шаблона отчета в системе административного мониторинга [Текст] / Фролов А.И., Лукьянов П.В. // Прикладная математика, управление и информатика. Сборник трудов Междунар. молодеж. конф., Белгород: ФГАОУ ВПО «Белгородский государственный национальный исследовательский университет», 2012, Т. 2. – сс. 175-178.
46. Ветров Д.П., Кропотов Д.А. Байесовские методы машинного обучения – М.: МГУ, 2007. – 132 с.
47. Паротькин Н.Ю. Автоматизация прогнозирования временных рядов генетическим алгоритмом // Молодёжь и наука: сборник материалов VII Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых, посвященной 50-летию первого полета человека в космос [Электронный ресурс] / отв. ред. О.А.Краев. Красноярск: Сиб. федер. ун-т., 2011. Режим доступа: [http:// http://conf.sfu-kras.ru/sites/mn2011/thesis/s6/s6_32.pdf](http://http://conf.sfu-kras.ru/sites/mn2011/thesis/s6/s6_32.pdf)
48. Фирстенко, А.Н. Метаописание временных рядов // Машинное обучение и анализ данных. 2011. №1. Т. 1. С. 111-112.
49. Лукьянов, П.В. К вопросу о прогнозировании состояния объекта мониторинга [Электронный ресурс] / Лукьянов П.В., Фролов А.И., С.И. Ашихмин // Информационные технологии в науке, образовании и производстве. ИТНОП-2012: материалы V-й Международной научно-технической конференции. – Режим доступа: <http://irsit.ru/files/article/223.pdf>. – 6 с.
50. Лукьянов, П.В. Некоторые аспекты поддержки принятия решений на основании данных административного мониторинга [Текст]/ А.В.

- Коськин, А.И. Фролов, П.В. Лукьянов, О.И. Аткищев // Информационно-измерительные и управляющие системы. – 2013. – №8. – С. 55-59
- 51.Голяндина, Н. Э. Метод “Гусеница”-SSA для анализа временных рядов с пропусками [Текст] / Голяндина Н. Э., Осипов Е. В. // Мат. модели. Теория и приложения. 2005. Вып. 6. С. 50-61.
- 52.Лукьянов П.В. Исследование характеристик выходных наборов данных системы административного мониторинга в контексте автоматизации прогнозирования состояния объекта мониторинга [Текст] / П.В. Лукьянов, А.И. Фролов // Вестник компьютерных и информационных технологий. – 2013. – № 1 – С. С. 17-21.
- 53.Чучуева И.А. Модель прогнозирования временных рядов по выборке максимального подобия: дис. ... канд. тех. наук. Московский Государственный Технический Университет им. Н.Э. Баумана, Москва, 2012.
- 54.Тихонов Э.Е. Прогнозирование в условиях рынка. Невинномысск, 2006. 221 с.
- 55.Jingfei Yang M. Sc. Power System Short-term Load Forecasting: Thesis for Ph.d degree. Germany, Darmstadt, Elektrotechnik und Informationstechnik der Technischen Universitat, 2006. 139 p.
- 56.Александров Ф. И., Голяндина Н. Э. Выбор параметров при автоматическом выделении трендовых и периодических составляющих временного ряда в рамках подхода «Гусеница»-SSA. СПб.: Петербургский гос. ун-т. 2005.
- 57.Леонтьева, Л.Н. Многомерная гусеница, выбор длины и числа компонент [Текст] / Л.Н. Леонтьева // Машинное обучение и анализ данных. – 2011. – №1. – Т. 1. – С. 2-10.
- 58.Голяндина, Н.Э. Метод «Гусеница»-SSA: анализ временных рядов: Учеб. Пособие. – СПб., 2004. – 76 с.

59. Резниченко, Е.В. Методы краткосрочного прогнозирования финансовых рынков [Текст] / Резниченко Е.В., Кочегурова Е.А. // Известия Томского политехнического университета. – 2007. – № 6 – Т. 1 – С. 19-23.
60. Box G.E.P., Jenkins G.M. Time series analysis: Forecasting and control. – San Francisco: Holden-Day, 1976. – 423 p.
61. Лукьянов, П.В. Анализ существующих подходов и моделей, обеспечивающих решение задач прогнозирования по результатам административного мониторинга [Электронный ресурс] / Фролов А.И., Лукьянов П.В. // Материалы II международной научно-технической интернет-конференции «Информационные системы и технологии», г. Орел, 01 апреля – 31 мая 2013 г. – Орел. – Режим доступа: <http://irsit.ru/files/article/341.pdf>
62. Фадеев, И.В. Метод SSA для прогнозирования временных рядов // Машинное обучение и анализ данных. 2011. №1. Т. 1. С. 109-110.
63. Abraham B., Ledolter J. Statistical methods for forecasting. – New York: Wiley, 2003. – 421 p.
64. Комиссарова, А.С. Сравнительный анализ мощности критериев проверки случайности и отсутствия тренда [Текст] / Комиссарова А.С., Лемешко Б.Ю. // Материалы Российской НТК “Информатика и проблемы телекоммуникаций”, Новосибирск. 2011. – Т.1. – С.72-75.
65. Беркович, А.С. Исследование распределений статистик критериев тренда и случайности [Текст] / Беркович А.С., Лемешко Б.Ю., Щеглов А.Е. // Материалы X международной конференции “Актуальные проблемы электронного приборостроения” АПЭП-2010. Т.6, Новосибирск, 2010. – С.13-17.
66. Кобзарь А. И. Прикладная математическая статистика. — М.: Физмат-лит, 2006.
67. Токмакова, А.А. Выделение периодической компоненты из временного ряда // Машинное обучение и анализ данных. 2011. №1. Т. 1. С. 31-41.

68. Садовникова, Н. А. Анализ временных рядов и прогнозирование. / Садовникова Н. А., Р.А. Шмойлова – М. : Изд-во Московского гос. Ун-та экономики, статистики и информатики, 2001. – 67 с.
69. Кельтон В., Лоу А. Имитационное моделирование. Классика CS. 3-е изд. – Спб.: Питер: Киев: Издательская группа BHV, 2004. – 847 с.: ил. ISBN 5-94732-981-7, ISBN 966-552-118-7.
70. Инмон, Б. DW 2.0: хранилища данных следующего поколения [Текст] : Открытые системы/ Б. Инмон. //Открытые системы. СУБД., 2007 - сс.20-26
71. Tropashko, V. Trees in SQL: Nested Sets and Materialized Path DBAzone.com [Электронный ресурс] / V. Tropashko. //DBAzone.com. - Режим доступа: [http://www/dbazine.com/oracle/or-articles/tropashko4](http://www.dbazine.com/oracle/or-articles/tropashko4)
72. Басс, Л. Архитектура программного обеспечения на практике [Текст] / Л. Басс, П. Клементс, Р. Кацман. - 2-е изд. - С-Пб. - 2006.
73. Фаулер, М. Архитектура корпоративных программных приложений [Текст] / М. Фаулер. - М.: Вильямс, 2007.
74. Якубайтис, А.Э. Теория автоматов [Текст]. Т. 13./ А.Э. Якубайтис, В.О. Васюкевич, А.Ю. Гобземис, Н.Е. Зазнова, А.А. Курмит, А.А. Лоренц, А.Ф. Петренко, В.П. Чапенко. - Москва: ВИНТИ, 1976. - сс.109-188
75. Шуршев, В.Ф. Опережающее управление социально-экономическими системами корпоративного типа [Текст]: АГТУ/ В.Ф. Шуршев, Н.П. Ганюкова. //Вестник АГТУ. Сер: Управление, вычислительная техника и информатика, 2010 - сс.23-27
76. Франчук, В.И. Основы построения организационных систем [Текст] / В.И. Франчук. - М.: Экономика, 1991. - с.454
77. фон Берталанфи, Л. Общая теория систем - обзор проблем и результатов [Текст] / Л. фон Берталанфи. //В ежегоднике "Системные исследования", 1969 - с.250
78. Флейшман, Б.С. Основы системологии [Текст] / Б.С. Флейшман. - М.: Радио и связь, 1982. - с.368

79. Урманцев, Ю.А. Общая теория систем: состояние, приложения и перспективы развития. "Система, симметрия, гармония" [Текст] / Ю.А.. - М.: Мысль, 1988. - сс.38-124
80. Силюк, Н.А. Организация управленческого труда [Текст] / Н.А. Силюк, П.В. Веселов, В.В. Галахов. - М.: Экономика, 1986. - с.5
81. Сетров, М.Н. Принцип системности и его основные понятия [Текст]: Мысль / М.Н. Сетров. // Проблемы методологии системного исследования, 1970 - сс.7-48
82. Рамин, М.Л. Региональное программно-целевое планирование и управление [Текст] / М.Л. Рамин, В.И. Мишин. - АВОТС, 1985. - с.9
83. Прангишвили, И.В. Системный подход и общесистемные закономерности [Текст] / И.В. Прангишвили, Н.А. Абрамова, В.Ф. Спиридонов. - М: СИНТЕГ, 2000. - с.528
84. Поспелов, Д.А. Ситуационное управление: теория и практика [Текст] / Д.А. Поспелов. - М.: Наука, 1986.
85. Перегудов, Ф.И. Основы системного анализа [Текст] / Ф.И. Перегудов, Ф.П. Тарасенко. - Томск: Изд-во науч.-техн. лит, 1997. - с.389
86. Охтилев, М.Ю. Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических объектов.[Текст] / М.Ю. Охтилев, Б.В. Соколов, Р.М. Юсупов - М.: Наука, 2006. - 410 с.
87. Оптнер, С.Л. Системный анализ для решения деловых и промышленных проблем, Пер. с англ. [Текст] / С.Л. Оптнер. - М.: Сов. радио, 1970. - с.216
88. Овчинников, В.Г. Методология проектирования автоматизированных информационных систем. Основы системного подхода [Текст] / В.Г. Овчинников. - М.: Компания Спутник +, 2005. - с.286
89. Соколов, Б.В., Цивирко Е.Г., Юсупов Р.М. Анализ влияния информатики и информационных технологий на развитие теории и систем управления сложными объектами [Текст] / Б.В. Соколов, Е.Г. Цивирко, Р.М. Юсупов // Труды СПИИРАН. - 2009. - Вып. 11. - С. 11-49

- 90.Соколов, Б.В. Комплексное моделирование функционирования автоматизированной системы управления навигационными космическими аппаратами [Текст] / Б.В. Соколов, Р.М. Юсупов // Проблемы управления и информатики. 2002. № 5. С. 24-41.
- 91.Моисеев, Н.Н. Теория ноосферы и математические модели [Текст] : Наука/ Н.Н. Моисеев. //Философия и социология науки и техники: Ежегодник, 1987 - сс.94-114
- 92.Соколов, Б.В Междисциплинарный подход к комплексному моделированию рисков при выработке управленческих решений в сложных организационно-технических системах [Текст] / Б.В. Соколов, Р.М. Юсупов // Сб. науч. тр. Междунар. науч. школы "Моделирование и анализ безопасности и риска в сложных системах (МАБР-2008)". - С. 146-155.
- 93.Мильнер, Б.З. Системный подход к организации управления [Текст] / Б.З. Мильнер, Л.И. Евенко, В.С. Раппопорт. - М.: Экономика, 1983. - с.336
- 94.Месарович, М. Теория иерархических многоуровневых систем [Текст] / М. Месарович, Д. Мако, И. Такахара. - Москва: Мир, 1973. - с.342
- 95.Йордан, Э. Структурные модели в объектно-ориентированном анализе и проектировании [Текст] / Э. Йордан, С. Аргила. - М.: Лори, 1999. - с.288
- 96.Ивахненко, А.Г. Долгосрочное прогнозирование и управление сложными системами [Текст] / А.Г. Ивахненко. - Киев: Техника, 1975. - с.311
- 97.Система мониторинга хода и результатов реализации ФЦП "Русский язык (2006-2010 годы)" / И.С. Константинов, А.И. Фролов, Н.А. Кравцова и др. - № 2009613809. - Зарегистрировано в Реестре программ для ЭВМ 16.07.2009.
- 98.Адаптивная система административного мониторинга. / И.С. Константинов, А.И. Фролов, Н.А. Кравцова и др. - № 2010611666. - Зарегистрировано в Реестре программ для ЭВМ 02.03.2010.

99. Коськин, А.В. Технологическая среда для комплексного сопровождения процессов информатизации организационно-технических систем. [Текст] - А.В. Коськин. - М: Машиностроение-1, 2006. - 230 с.
100. Коськин, А.В. Управление производственно-образовательными комплексами на основе информационно-аналитических ресурсов: автореферат. [Текст] - А.В. Коськин. - СПб.: СПбГТИ(ТУ), 2007. - 16 с.
101. Когут, А.Е. Информационные основы регионального социально-экономического мониторинга [Текст] / А.Е. Когут, В.С. Рохчин. - С-Пб: ИСЭП РАН, 1995. - с.141
102. Бодрунов, С.Д. Проблемы, принципы и методы корпоратизации авиапромышленного комплекса России [Текст] / С.Д. Бодрунов, О.Н. Дмитриев, Ю.А. Ковальков, Д.В. Мантуров. - Сп-Б.: Корпорация "Аэрокосмическое оборудование" и "Петроградский и К°", 2000. - с.432
103. Константинов, И.С. Экологическая безопасность промышленно-транспортного комплекса как объект управления АСУ [Текст] / И.С. Константинов, О.А. Иващук // Информационные технологии в науке, образовании и производстве. ИТНОП-2010: материалы IV-й Международной научно-технической конференции, 2010. - Т. 1. - С. 25-35
104. Константинов, И.С. Автоматизированная система управления экологической безопасностью промышленно-транспортного комплекса. [Текст] / И.С. Константинов, О.А. Иващук // Вестник компьютерных и информационных технологий. - 2009. - № 8. - С. 44-49.
105. Константинов, И.С. Адаптивное управление экологической безопасностью промышленно-транспортного комплекса. [Текст]/ И.С. Константинов, О.А. Иващук // Научные Ведомости Белгородского государственного университета. - 2009. - № 7(62)2009. - Вып. 10 - С. 53-58.
106. Фролов, А.И. Особенности подготовки процедуры сбора данных при использовании автоматизированной адаптивной системы административного мониторинга / А.И. Фролов, О.В. Матюхина //

- Информационные системы и технологии: материалы Международной научно-технической Интернет-конференции. Орел ФГОУ ВПО "Государственный университет-УНПК", 2011. С. 140-144.
107. Волков, В.Н. Исследование информационных потоков в системе управления промышленным предприятием Текст. // Сборник трудов "Известия ОрелГТУ. Информационные технологии". Орел: изд-во Орл. гос. ун-та, 2005. - Выпуск 2 (8).- С. 6-9.
108. Волков, В.Н. Информационное обеспечение системы организационного проектирования для промышленных предприятий [Текст] / Волков В.Н., Савина О.А. // "Информационные технологии в науке, образовании и производстве". Материалы международной научно-технической конференции. - Орел: ОрелГТУ, 2006, Т4. - С.18-23.
109. Открытые стандарты для построения Workflow-систем [Электронный ресурс] // Workflow Management Coalition. - Режим доступа: <http://www.wfmc.org>
110. Szyperski, C. Component Software - Beyond Object-Oriented Programming [Текст] / C. Szyperski. - Addison-Wesley, 2002. - с.608
111. Somasegar, S. Руководство Microsoft по проектированию архитектуры приложений [Текст] / S. Somasegar, S. Guthrie, D. Hill. - Microsoft, 2009.
112. Royce, W. Managing the Development of Large Software Systems [Текст] / W. Royce. - IEEE Wescon, 1970. - сс.1-9
113. Parsaye, K. Surveying Decision Support: New Realms of Analysis [Текст]: Database Programming and Design/ K. Parsaye., 1996 - №4. - сс.35-48
114. Parsaye, K. Characterization of Data Mining Technologies and Processes [Текст] : The Journal of Data Warehousing/ K. Parsaye., 1998 - №1. - сс.2-15
115. White, S.A. Process Modeling Notations and Workflow Patterns [Текст]: Workflow Handbook/ S.A. White. - IBM Corp, 2007.
116. Minto, B. The pyramid principle: logic in writing and thinking [Текст] - Financial Times Prentice Hall, 2008. - с.177

117. Kolliat, G. OLAP, relational, and multidimensional database systems [Текст]: ACM SIGMOD. Т. 25./ G. Kolliat. - New York: ACM, 1996. - сс.64-69
118. Dahl, O.J. Structured programming [Текст] / O.J. Dahl, E.W. Dijkstra, C.R. Hoare. - London; New York: Academic Press, 1972. - с.220
119. Cormen, T.H. Cormen, T.H. Introduction to Algorithms, Second Edition [Текст] / T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein. - MIT Press and McGraw-Hill, 2001. - сс.527-529
120. OpenLaszlo Showcase [Электронный ресурс] //OpenLaszlo. - 2011. - Режим доступа: <http://www.openlaszlo.com/node/409>
121. OASIS Standard. OASIS //OASIS. Web Services Business Process Execution Language (WS-BPEL), Version 2.0. - 2008. - Режим доступа: <http://docs.oasis-open.org/wsbpel/2.0/0s/wsbpel-v2.0-0s.pdf>
122. Manual of Style for Technical Publications [Текст]: Microsoft Corporation Editorial Style Board - 3-е изд. - Microsoft, 2004. - с.432
123. Information technology. Open Distributed Processing. Unified Modeling Language (UML) Version 1.4.2. [Текст]: ISO/IEC 19501:2005 – Switzerland.